

Transformer 深層学習モデルを用いた人間の動作における潜在表現の獲得と別部位データの生成

研究代表者 島 孔 介 名古屋工業大学 大学院工学研究科・工学専攻 助教

1 研究背景

近年、慣性計測装置（IMU）を用いたモーションキャプチャは、スポーツ解析やリハビリテーション、技能評価など様々な分野で利用されている。一方、全身運動を高精度に取得するためには、多数のセンサを身体各部位へ装着する必要がある、装着負担や計測コストが大きな課題となっている。

本研究では、「身体各部位の運動には潜在的な連動性が存在する」という仮説に基づき、少数の IMU から他部位の運動を生成する深層学習モデルの実現を目的とする。

2 提案手法

本研究では、多変量時系列予測モデルとして提案されている iTransformer を、異なる身体部位間の時系列データ変換へ適用可能な Encoder-Decoder モデル「iTransposer」（図 1）へ拡張した。

従来の iTransformer は、将来時刻の時系列予測（Forecasting）を目的として設計されており、Encoder のみから構成される。一方、本研究で対象とする部位間変換は、入力系列と出力系列が異なる「翻訳」に近い問題である。例えば、腰部 IMU を入力として右手首 IMU を生成する場合、入力系列と出力系列は異なる物理量であり、将来予測とは異なる特徴を持つ。そのため、本研究では Transformer 本来の Encoder-Decoder 構造を導入し、入力部位と出力部位の関係を学習可能なモデルを構築した。

Encoder では、入力部位の各チャンネル（3 軸加速度、加速度・角速度・地磁気など）をトークンとして扱い、多変量時系列に対する Self-Attention により部位内およびチャンネル間の関係を学習する。各 Transformer Block は、Multivariate Attention, Feed-forward Network, Layer Normalization, Residual Connection から構成される。

Decoder では、Encoder が獲得した潜在表現を Cross Attention により参照しながら、変換先部位の時系列を生成する。Decoder においても Multivariate Attention を採用し、Encoder が抽出した入力部位の特徴と、出力部位の自己相関を同時に考慮した特徴生成を行う。

また、Decoder の入力方法についても検討を行った。当初は BERT に類似した Masked Target を入力とする学習を行ったが、推論結果が入力部位の波形に引き寄せられる傾向が見られた。そこで、Transformer の翻訳モデルで一般的に用いられる Shifted Target を Decoder 入力として採用したところ、変換精度が改善することを確認した。このことから、部位間変換においては、変換先系列の時系列連続性を Decoder に与えることが有効であることが示唆された。

さらに、本研究ではデータセットに依存しない学習基盤を構築した。前処理では、各データセットに対して共通の正規化、スライディングウィンドウ化、データセット生成を行うことで、テコンドーデータおよび歩行オープンデータセット COMPWALK-ACL の双方に対して、同一の学習・推論コードを適用可能とした。

現段階では、Pelvis（骨盤）から Right Hand（右手）のような単一部位変換を対象としているが、将来的には Encoder 入力チャンネル数と Decoder 出力チャンネル数を独立に扱えるようモデルを拡張し、少数センサから全身 22 セグメント（66 チャンネル）の運動を同時生成することを目標としている。

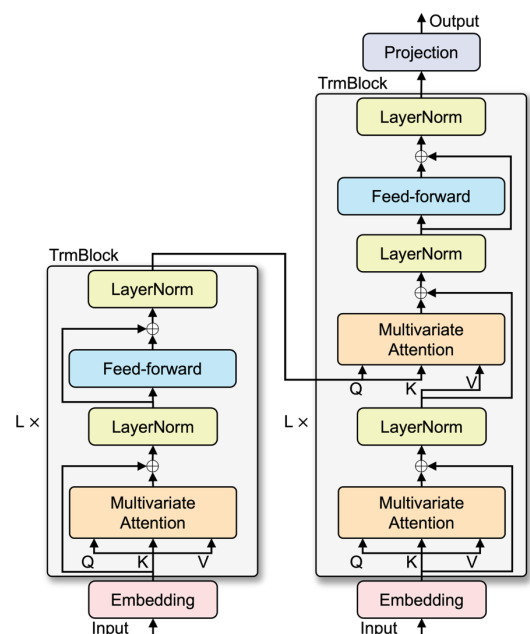


図 1 iTransposer アーキテクチャ

3 テコンドーデータによる検証

岐阜テコンドークラブの協力のもと収集した IMU データを用いて、腰、右手首、左手首、右足首、左の 5 部位のデータセットを構築した。

腰部 IMU を入力、右手首 IMU を出力とする変換タスクを例題とし、ハイパーパラメータ探索を行った結果、提案モデルは入力データをそのままコピーするのではなく、変換先の時系列特徴を学習し、右手首の運動を再構成できることを確認した（図 2）。

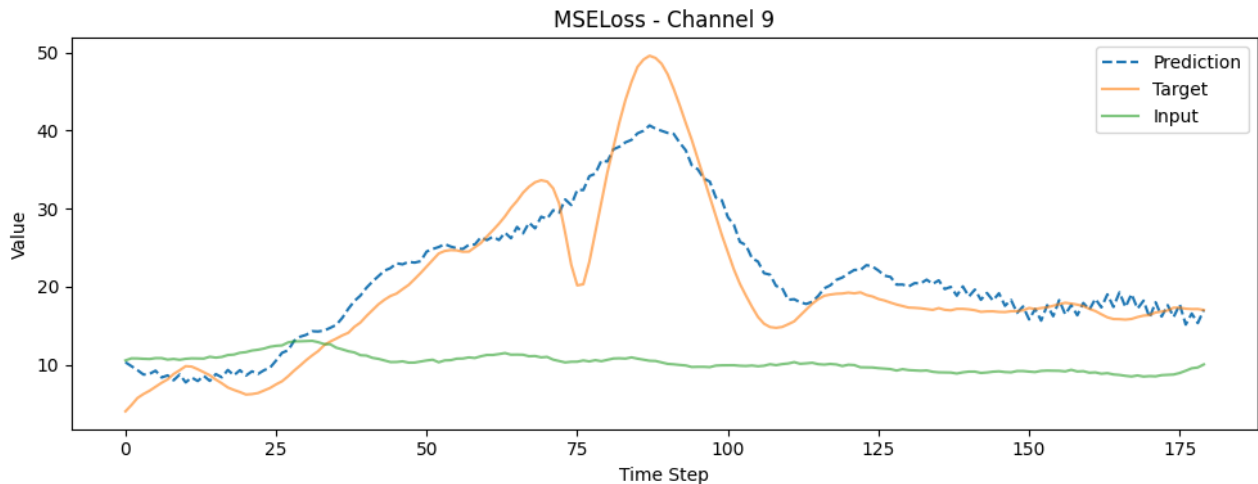


図 2 腰データ→右手首データ変換結果例

4 歩行オープンデータセットへの展開

提案モデルの汎用性を検証するため、新たに歩行オープンデータセット COMPWALK-ACL を用いた評価を行った。COMPWALK-ACL は、骨盤・体幹・頭部・上肢・下肢を含む 23 身体セグメントの加速度データを 100 Hz で計測した大規模データセットであり、健康者および ACL 損傷者を含む歩行データが公開されている。本研究では、このデータセットを利用することで、独自に収集したテコンドーデータとは異なる運動様式に対しても提案モデルが適用可能であることを検証した。

まず、COMPWALK-ACL に含まれる Segment Acceleration データを対象とし、学習用データセットを新たに構築した。データセット構築では、各試行を一定長のスライディングウィンドウへ分割し、入力系列・出力系列を正規化した上で学習データとして保存する前処理パイプラインを実装した。また、従来はテコンドーデータ専用となっていた前処理コードを整理し、データセットに依存しない共通のデータ読み込み・正規化・ウィンドウ化処理へと改良した。その結果、テコンドーデータおよび COMPWALK-ACL の双方に対して、同一の学習コードを用いることが可能となった。

モデルの初期検証として、Pelvis（骨盤）の 3 軸加速度を入力とし、Right Hand（右手）の 3 軸加速度を生成する部位間変換タスクを構築した。これは、将来的な「少数センサから全身運動を生成する」研究へ向けた第一段階として、まず単一部位間の変換性能を評価することを目的としている。

さらに、Decoder 入力方式についても検討を行った。当初は BERT 型モデルを参考に、変換先データを部分的にマスクした Masked Target を Decoder 入力として学習を行った。しかし、推論結果には入力部位（Pelvis）の波形へ引き寄せられる傾向が見られ、部位間変換として十分な性能は得られなかった。そこで、Transformer の翻訳モデルで一般的に用いられる Shifted Target を Decoder 入力として採用したところ、学習損失が改善するとともに、生成波形も変換先部位の特徴へ追従する傾向が確認された（図 3）。特に Right Hand の一部チャンネルでは、入力波形ではなく変換先波形を追従する様子が確認され、Decoder に時系列連続性を与えることの有効性が示唆された。

一方で、生成結果はチャンネルごとに精度の差が見られ、入力部位の特徴を強く保持するチャンネルも存在した。このことから、部位間変換には単純な時系列生成だけではなく、身体各部位間の時間的・空間的な連動性をより適切に表現する潜在表現の獲得が重要であると考えられる。

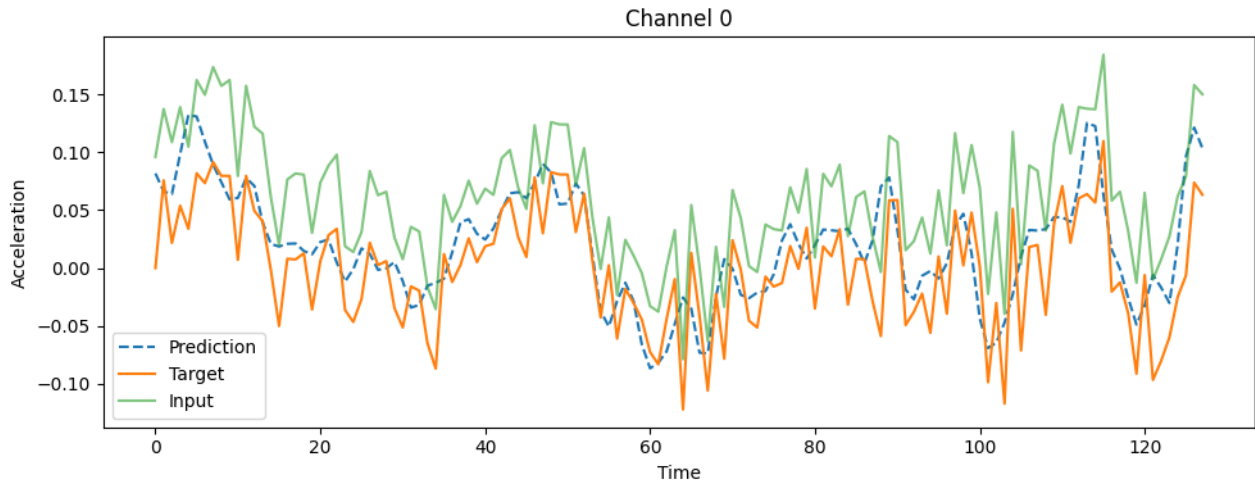


図 3 歩行オープンデータを用いた腰→右手首データ変換結果例

今後は、現在の単一部位変換を基礎として、Pelvis の少数センサ入力から全 22 身体セグメント（66 チャンネル）の運動を同時生成可能なモデルへ拡張し、少数センサによる全身運動再構成の実現を目指す。

5 今後の展望

本研究では、テコンドーデータおよび歩行オープンデータセット COMPWALK-ACL を用いて、少数センサから別部位の運動を生成する部位間変換モデル iTransposer を構築し、腰部から右手部への変換が可能であることを確認した。また、Decoder 入力方式を検討した結果、Shifted Target を用いることで変換精度が改善することを確認し、部位間変換タスクにおける Decoder 設計の重要性を明らかにした。

一方、現在のモデルは Encoder 入力チャンネル数と Decoder 出力チャンネル数が一致することを前提としているため、Pelvis（3 チャンネル）から Right Hand（3 チャンネル）のような単一部位変換を対象としている。しかし、本研究の最終的な目的は、少数の IMU のみを装着した状態から、全身運動を推定することである。

そのため今後は、Encoder 入力チャンネル数と Decoder 出力チャンネル数を独立に扱えるようモデル構造を拡張し、Pelvis の少数チャンネル入力から、全 22 身体セグメント（66 チャンネル）の運動を同時生成可能な iTransposer を開発する予定である。これにより、従来の時系列予測モデルでは扱うことのできなかった「異なるチャンネル数間の時系列変換」を実現し、身体各部位間の連動性をより適切に表現する潜在表現の獲得を目指す。

さらに、将来的には歩行だけでなく、スポーツ動作や日常生活動作など、多様な身体運動に対して評価を行い、少数センサによる全身運動再構成モデルとしての有効性を検証する予定である。本研究が実現されれば、モーションキャプチャに必要なセンサ数を大幅に削減できるだけでなく、スポーツ技能評価、リハビリテーション支援、ヒューマンインタフェースなどへの応用が期待される。

【参考文献】

- VASWANI, Ashish, et al. Attention is all you need. Advances in neural information processing systems, 2017, 30.
- LIU, Yong, et al. itransformer: Inverted transformers are effective for time series forecasting. In: International conference on learning representations. 2024. p. 11116-11140.
- Yona, T., Peskin, B. & Fischer, A. The COMPWALK-ACL: A Dataset of Multi-pace IMU Gait Kinematics in Adolescents, Adults, and ACL Injured Patients. Sci Data 13, 4 (2026).

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
動的社会ネットワーク可視化のためのノードの質量を考慮したグラフ描画アルゴリズム	情報処理学会論文誌数理モデル化と応用 (TOM), 18 巻, 3 号, p. 109-118, 発行日 2025-08-28	2025 年 8 月 28 日