

SubgraphX を用いた ST-GCN のアテンション分析による 組み立て作業の熟練動作分析手法

A method for analyzing skilled movement of assembly tasks
by attention analysis of ST-GCN using SubgraphX.

棚橋悠翔¹ 木村康佑² 秋月秀一² 橋本学² 岡本雅己³

Yuto Tanahashi¹, Kosuke Kimura², Shuichi Akizuki², Manabu Hashimoto², and Masaki Okamoto³

¹ 中京大学 工学部

¹School of Engineering, Chukyo University

² 中京大学大学院 工学研究科

²Graduate School of Engineering, Chukyo University

³ 日産自動車株式会社

³Nissan Motor Corporation

Abstract: In the manufacturing field, automatic analysis of actions is needed from the viewpoint of improving the efficiency of worker training. Conventional methods create machine learning models to identify the degree of skill and visualize the basis for judgment, but these methods have problems in that it is difficult to identify the degree of skill for unknown workers, and spatial analysis is not possible. In this study, we improved these problems by adding differential processing to the machine learning model and using SubgraphX for analysis. We confirmed that the identification rate was improved by 29% and that the analysis results were consistent with visual analysis.

1. はじめに

製造現場では、熟練作業者の不足が問題視されており、教育効率化の観点から、レベルに応じた作業者への適切な改善案フィードバックが求められている[1]。このためには、個別の動作に関して改善すべき箇所を自動分析する必要があり、関連研究として、動作計測データをもとに技能レベルを識別できる機械学習モデルを構築し、この判断根拠を可視化する手法がある[2]。この手法は、作業者のスケルトンデータを直接的に用いているため、簡便である一方で、学習データに含まれる個々の作業者特有の情報に特化して学習してしまい、学習データに含まれない未知の作業者の技能レベル識別が難しくなるという問題点がある。例えば、今回使用した学習データには作業者の骨格の長さの情報が含まれているが、これは熟練度とは関係がない、作業者によって異なる作業者特有の情報であり、熟練度識別に不要な情報を学習させてしまうため悪影響を及ぼしていると考えられる。また、分析結果についても従来法においては時間的な分析のみであり、体のどの部位かといった空間的な分析ができないという問題がある。

本研究では、技能レベルを識別できる機械学習モ

デルとして、スケルトンデータの直接的な入力可能な ST-GCN[3]を採用し、この入力としてデータの微分処理を追加することによって、技能レベル識別の問題を改善する。また、時間的かつ空間的な分析のため、グラフニューラルネットワークのアテンションの分析手法のうち、サーベイ論文[4]の比較実験において最も説明性が高く、出力形式の解釈性が高い SubgraphX[5]を機械学習モデルの判断根拠の可視化に適用することによって、空間的な分析ができない問題を改善する。

2. 提案手法

本章では、熟練度識別の高精度化手法と SubgraphX によるアテンション分析手法の 2 種類の提案手法について説明する。以下、2.1 節では熟練度識別の高精度化手法について述べ、2.2 節では SubgraphX による時空間アテンション分析手法について述べる。

2.1. 微分処理による熟練度識別の高精度化

従来の分析手法において、学習に用いている作業者のスケルトンデータには、図 1 に示すような「動き」「立ち位置」「骨格の長さ」の 3 種類の固有情報

が含まれていると考えられる。この内、熟練度識別において学習したい情報は「動き」の情報であり、「立ち位置」と「骨格の長さ」のような初期値の情報は不要であると考えられる。そこで、これらの初期値情報を削減しつつ、動きの情報を保持する前処理を加えることで熟練度識別の高精度化が可能であると考えた。本研究では、初期値情報を削減しつつ、動きの情報を保持する前処理として、スケルトンデータの微分情報への変換を施すこととした。

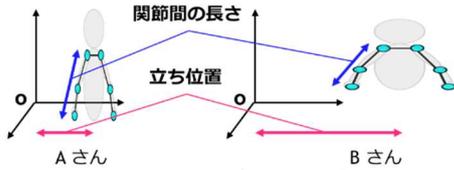


図1 スケルトンデータに含まれる作業者特有の情報

提案手法を図2に示す。提案手法では、オリジナル ST-GCN モデル[3]の前段に微分処理を追加し、速度データを入力する。これにより、図1に示すような個々の作業者特有の情報が低減されることを期待している。

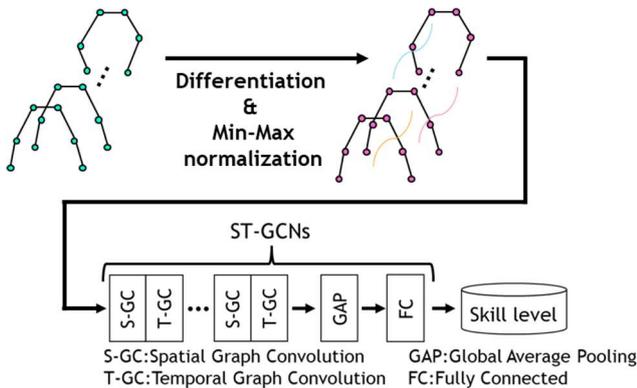


図2 速度入力型 ST-GCN

2.2. SubgraphX によるアテンション分析

提案手法を図3に示す。従来法における機械学習モデルの判断根拠の可視化において、SubgraphX を適用し、アテンションを分析する。これにより、時間的かつ空間的な分析が可能になることを期待している。

SubgraphX とは、グラフデータを扱うネットワークにおいて、入力されたグラフデータから識別に重要であったサブグラフを特定する手法である。

SubgraphX のアルゴリズムとしては、まず入力データにマスクを加えたサブグラフを作成し、モデルに入力する。次に、モデルの出力値からサブグラフ

を評価する。この際、サブグラフの評価には、モデルへの寄与率を調査する Shapley 値を用いる。そして、これらの2つの処理を何度も実行し、最終的に評価値が最大のサブグラフを返すというアルゴリズムである。この際、評価値が最大となるサブグラフを効率的に探索するため、モンテカルロ木探索をおこなう。1回のモンテカルロ木探索における SubgraphX の処理イメージを図4に示す。

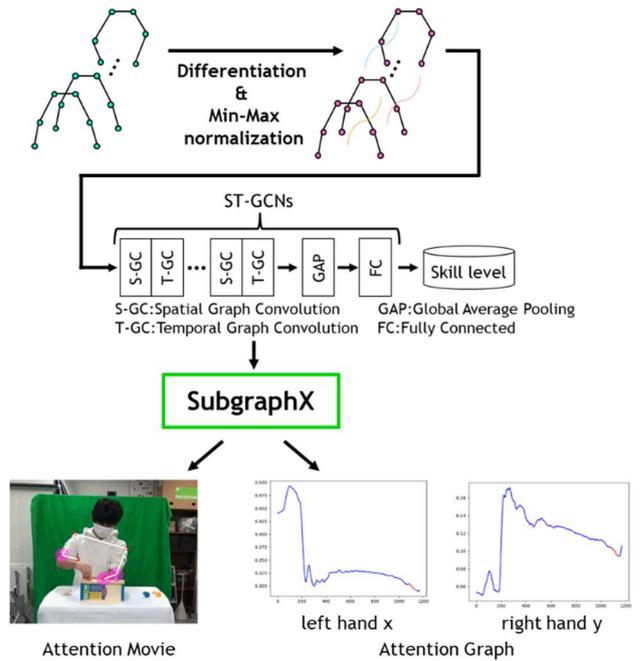


図3 SubgraphX による時空間アテンション分析

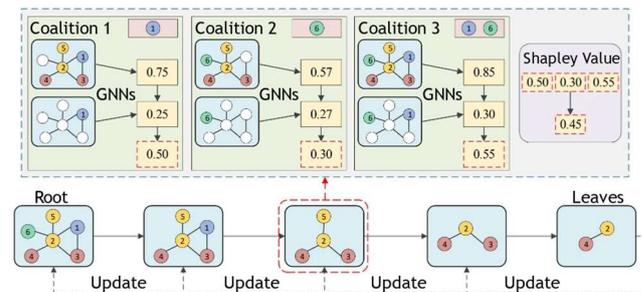


図4 SubgraphX における1回のモンテカルロ木探索のイメージ

SubgraphX のマスク方法については図5に示す 0feature と split の2種類の実装方法がある。

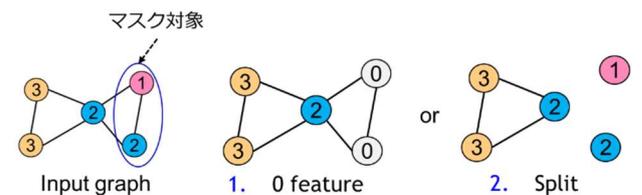


図5 SubgraphX のマスク方法

Ofeature ではマスク対象となるノードの特徴量を 0 にすることでマスクを表現するが、split においてはマスク対象のノードとのエッジの接続を切ることでマスクを表現するという違いがある。本実験においては、プログラムの実装負荷の観点から、実装が容易である Ofeature を採用した。

3. 実験と考察

本研究では、ボルト 3 本をレンチにより取り付ける組立作業のデータを、被験者 14 人に対して各 90 本、合計 1260 本取得した。組み立て作業に用いた対象物を以下の図 6 に、実際の組み立て動作の例と取得したデータの例を以下の図 7 に示す。

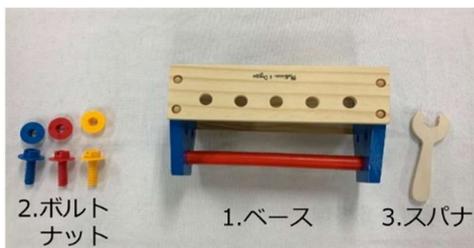


図 6 組み立て作業に用いた対象物



図 7 組み立て動作の例と取得したデータの例

このデータセットを用いて、以下の 2 つの実験をおこなった。

3.1. 熟練度識別実験

本実験では、オリジナルの ST-GCN と、提案手法を用いて技能レベル識別をおこない、性能を比較した。結果を表 1 に示す。識別性能は識別率と F 値で表現した。提案手法では識別率 91%、F 値が 0.918 であり、従来法より性能が大幅に改善したことを確認した。

表 1 実験結果

	Accuracy	F1-score
ST-GCN (従来法)	62%	0.629
速度入力型 ST-GCN(提案手法)	91%	0.918

この結果から、微分処理によって速度情報を入力することが、学習データに含まれる個々の作業者特有の情報の影響を低減し、識別性能の改善に繋がったと考えられる。特に、図 1 に示すような作業者の立ち位置の情報と関節間の長さの情報は座標データの絶対位置に関係があり、微分処理によって相対的な速度データにしたことで影響が低減されたのではないかと推察される。

3.2. SubgraphX によるアテンション分析

本節では、SubgraphX によるアテンション分析に関連しておこなった 3 つの実験について述べる。

はじめに、入力データのシーケンス長と実行時間の関係を調査した。今回用いた SubgraphX は他のグラフニューラルネットワークの分析手法に比べて、説明性が高い分、探索に時間がかかるという問題があった。そこで、本実験で用いたデータのうち、シーケンス長が最短のデータで調査したところ、以下の表 2 のような実行時間となった。シーケンス長 104, 52, 26 についてはそれぞれ元のシーケンス長の 1/16, 1/32, 1/64 の長さである。

表 2 実行時間の調査結果

シーケンス長	1662	104	52	26
実行時間	1 週間以上	1 週間以上	1.5-2 時間	10-20 分

調査の結果、シーケンス長が 104 シーケンス以上の場合には実行時間が 1 週間以上かかり、現実的でないことが明らかになった。また、52 シーケンスの場合には作業時間が長いデータの場合に 1fps 程度のデータとなり、動作の内容や特徴が欠落するため、分析や識別に不適切なデータとなる可能性があると考えられる。そのため、これ以降の実験は、組み立て作業のうち、ねじの把持、ねじの位置合わせ、ねじの仮締めのための短い作業を用いて実験をおこなった。また、実行時間の問題から、シーケンス長を 52 にダウンサンプリングし分析に用いた。

次に、時間方向と空間方向におけるアテンションの分布について実験をおこなった。具体的には、初心者のアテンション動画 58 本のうち 9 本をランダムサンプリングし、表出したアテンションの確率分布を調査した。時間方向のアテンションの確率分布を図 8 (a) に、動作 f における空間方向のアテンションの確率分布を図 8 (b) に示す。

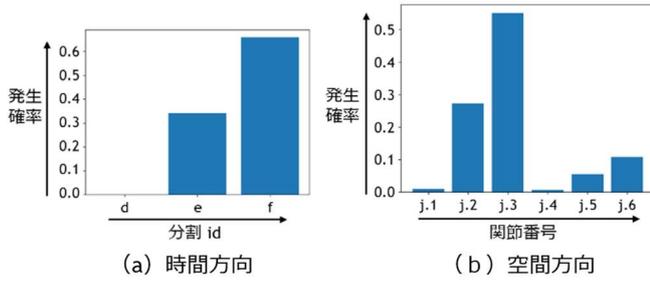


図 8 各方向のアテンションの確率分布

時間方向のアテンションの分布については横軸がそれぞれ組み立て作業中の特定の動作を表しており、d がねじの把持、e がねじの位置合わせ、f がねじの仮締めである。また、縦軸には各アテンションの発生確率を示す。この図から、ST-GCN が動作 f を注視する傾向があることが明らかになった。そこで、動作 f にアテンションが発生したデータに限定し、空間方向のアテンションを調査した。

空間方向のアテンションの分布についてはそれぞれが特定の関節番号に対応しており、j.1 は右肩、j.2 は右肘、j.3 は右手首、j.4 は左肩、j.5 は左肘、j.6 は左手首を示している。また、縦軸には各アテンションの発生確率を示す。この図から、動作 f においては右手首と右肘にアテンションが集中していることが明らかになった。また、左手首と左肘についてもわずかにアテンションが集中することが明らかになった。

最後に、2 つ目の実験で明らかになった、初心者データのアテンションが集中している 2 箇所について、熟練者の作業動画との比較を目視でおこなった。その結果、以下の図 9、図 10 に示す 2 種類の動作の差異が明らかになった。



図 9 仮締め動作における右手の動作の差異



図 10 仮締め動作における左手の動作の差異

図 9 は仮締め動作における右手の動作の差異であり、ほとんどの動画で表出していた。また、図 10 は仮締め動作における左腕の動作の差異であり、10 本程度の動画で表出していた。これらの結果から、ST-GCN が注視した領域が、動作の差異が発生した部分と一致することが明らかになった。また、アテンションの発生確率が高く、より ST-GCN が注視していた領域ほど顕著に動作の差異が発生することが明らかになった。

4. まとめ

本研究では、技能レベルの識別モデル ST-GCN に微分処理を追加することによって、未知の作業者に対しても、F 値が 0.918 であり、オリジナルの ST-GCN より性能が向上したことを確認した。

また、SubgraphX による時空間アテンション分析により、アテンションの発生確率が高い領域ほど顕著に動作の差異が発生することを確認した。

注：本研究で用いたデータセットは、実験参加者からインフォームドコンセントを得たうえで取得したものである。

謝辞

本研究の一部は、JSPS 科研費 C (21K03984) の助成を受けて遂行されたものである。

参考文献

- [1] 澤武一, 伊藤伸英, 天野高宏, 横山淳一, 長谷川文昭, 寺田貢紀, 高野知行, “映像を用いた技能者教育の試みと効果 (第 2 報)”, Journal of JSEE, Vol. 59, No. 6, pp. 6_91-6_96, (2011)
- [2] R. Miyoshi, K. Kimura, S. Akizuki, M. Hashimoto, “Analysis of Human’s Skilled Process of Assembly Task using Time-sequence based Machine Learning”, International Workshop on Advanced Imaging Technology, Vol. 12177, pp. 241-246, (2022)
- [3] S. Yan, Y. Xiong and D. Lin, “Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition”, AAAI, Vol. 32, No. 1, (2018)
- [4] H. Yuan, H. Yu, S. Gui and S. Ji, “Explainability in Graph Neural Networks: A Taxonomic Survey”, IEEE Transactions on Pattern Analysis & Machine Intelligence, Vol. 45, No. 05, pp. 5782-5799, (2023)
- [5] H. Yuan, H. Yu, J. Wang, K. Li, and S. Ji, “On Explainability of Graph Neural Networks via Subgraph Explorations”, Proceedings of Machine Learning Research, Vol. 139, pp. 12241-12252, (2021)