CF-013

特徴量選択付き時間畳み込みネットワークを用いた作業行動の分節化

Operational Work Segmentation by Temporal Convolutional Network with Elastic Feature Selection

清水 裕斗[1)]　山本 泰生[1)]　西村 雅史[1)]
Yuto Shimizu　Yoshitaka Yamamoto　Masafumi Nishimura

塩野 由紀[2)]　白澤 怜樹[2)]　中野 貴行[2)]　青木 崇浩[1)2)]
Yuki Shiono　Reiki Shirasawa　Takayuki Nakano　Takahiro Aoki

1　はじめに

製造現場のデジタル化に伴い，複雑多様な人間の行動を定量的に分析・評価をしようとする動きが強まっている [1, 2]．**行動分節化** (*Action Segmentation*，以下，AS と略す) は作業標準化と作業保証を実現するための要素技術であり，行動を表す時系列データを基本動作単位に区分化するタスクである．しかし，一般的な AS モデルでは，推論の際にモデルがどの特徴量に注目して区分したのかを理解するために直接利用することはできない．

このようなモデルにとって重要な特徴量を可視化することは，対象とする作業の分析や熟練者の技の定量的な分析などに大きく貢献できる可能性があるため，非常に重要である．この課題の解決を目指した研究として，特徴量選択を伴う行動分節化モデルに関する研究 [13] が挙げられる．**特徴量選択** (*Feature Selection*) は機械学習のモデルを使用する際に有効な特徴量の組み合わせを探索する技術であり，使用する特徴量を少なくすることで解釈性を上げることができるといったメリットがある．先行研究 [13] のモデルは特徴量を選択するための重みフィルタを持ち，各工程に対応する重みフィルタを可視化することで，工程ごとにどの特徴量に注目して工程分解を行なっているかを特定する．しかし，先行研究 [13] のモデルは特徴量利用の節減効果の面に課題がある．本研究では，先行研究 [13] の課題を**重みフィルタの作成方法**，**学習方法**の 2 つの点で改善を行い，より高精度なモデルの提案を目指す．

2　関連研究

コンピュータビジョン分野において，AS を扱う様々な機械学習モデルの研究が行われている．教師あり学習の枠組みとして，TCN [4] ベースの手法が多く提案されている．MS-TCN[5] は，多段の膨張畳込層 (*Dilated Convolution layer*) から構成される TCN ユニットの出力を次の TCN ユニットの入力として多段階組み上げた構成をとる．広域受容野からの大域特徴を用いることで比較的安定した性能を得ている．MS-TCN の 1 段目を改良した MS-TCN++[6] が提案されている．また，行動を区分するモジュールの他に，行動が切り替わる境界を回帰予測するモジュール導入した手法が提案されている [10]．近年では，Transformer ベースの手法として ASformer[7] や UVAST[8] が提案されている．これらの行動分節化モデルの製造現場での行動解析への有効性について検証した研究として [14] が挙げられる．

1）静岡大学大学院総合科学技術研究科. Department of Computer Science, Shizuoka University.
2）ヤマハ発動機株式会社生産技術部. Manufacturing Technology Div., Yamaha Motor Co., Ltd.

特徴量選択は，大きく Filter method，Wrapper method，Embedded method に分類される [11]．Lasso[12] は強力な Embedded method の特徴量選択手法として知られている．Lasso は損失関数に L1 正則化項を持ち，係数がゼロになるようなスパース解を容易に取得できるという特性を持つ．そのため，Self-Attention[9] などの手法と比較した際に，モデルが対象の特徴量を使っているかどうかが一目で分かるため解釈が容易であるといったメリットがある．近年では，LassoLayer と呼ばれる特徴量選択を行う手法が提案されている [3]．LassoLayer は入力ベクトルと重み付けされた出力ベクトルの間が 1 対 1 で接続された 2 つの層で構成されている．LassoLayer の重みは L1 正則化によって学習されるため，重要度の低い入力に対応する重みはゼロとなるように縮小される．そのため，推論時には重要な特徴量のみを以降の層に渡すことが可能となる．

3　特徴量選択を伴う行動分節化モデル

本章では，先行研究 [13] で提案された特徴量選択を伴う行動分節化の手法を概説する．

3.1　手法の概要

図 1に先行研究 [13] の構造を示す．先行研究 [13] の手法は，3 つのモジュールから構成される．LassoLayer は特徴量選択を行うモジュールであり，工程ごとに異なる特徴量を使用するためのフィルタの役割を持つ．選択モジュールは LassoLayer の重みを選択するためのモジュールである．区分化モジュールは LassoLayer によって特徴量選択された入力を受け取り工程分解を行うモジュールである．選択モジュールおよび区分化モジュールは AS モデルによって構成される．

次に，t フレーム目の入力に対する特徴量選択および予測の流れを説明する．ここで，工程数 C，1 フレームあたりの特徴量数 N に対して LassoLayer の重みは $W = [\boldsymbol{w}_1, ..., \boldsymbol{w}_C]$ で表され，$\boldsymbol{w}_i (1 \leq i \leq C) \in \mathbb{R}^N$ となる．まず，t 番目の入力フレーム $\boldsymbol{x}_t \in \mathbb{R}^N$ に対して，選択モジュールにより使用する重みフィルタ $\hat{\boldsymbol{w}}_t$ を選択する．ここで，$\hat{\boldsymbol{w}}_t = \boldsymbol{w}_{\hat{z}_t}$ であり，$\hat{z}_t$ は式 (1) によって求められる．

$$\hat{z}_t = Max(\boldsymbol{z}_t) \tag{1}$$

$\boldsymbol{z}_t \in \mathbb{R}^C$ は t フレーム目の入力に対する選択モジュールの予測尤度であり，Max によって尤度が最大である工程番号を取り出す．つまり，$Max(\boldsymbol{z}_t)$ は t フレームに対応する工程番号と同一視できる．予測結果に応じて使用する重み $\hat{\boldsymbol{w}}_t$ の選択する．これにより，各工程に適したフィルタを選択できる．次に，選択した重み $\hat{\boldsymbol{w}}_t$ を使用して式 (2) の演算を行うことで，入力 x_t に対する出力 y_t

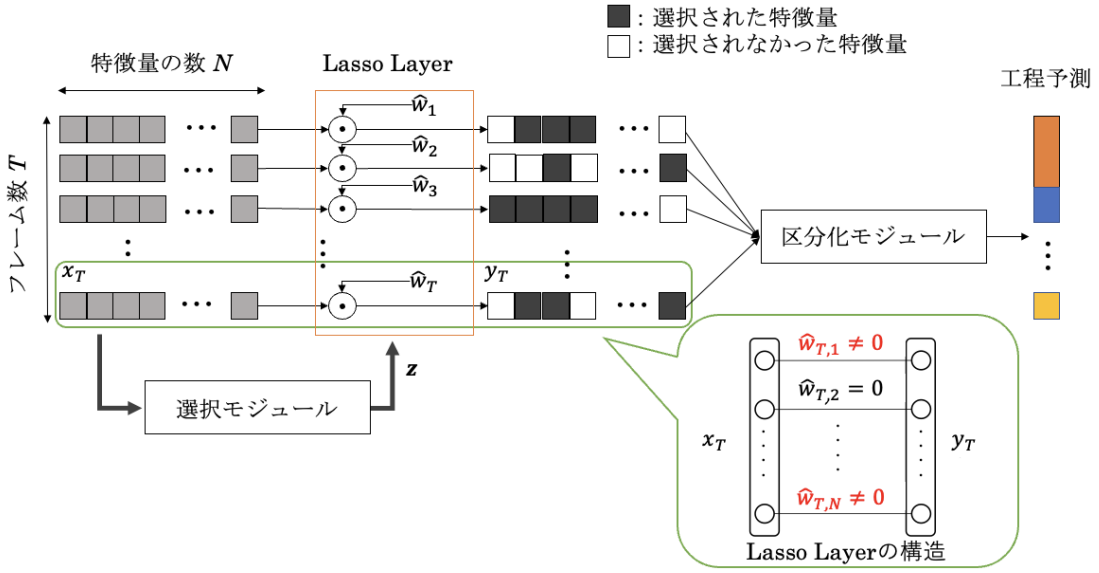

図 1　先行研究 [13] の手法の構造

を得る．ここで，⊙ は 2 つのベクトルのアダマール積を表す．

$$\boldsymbol{y}_t = \boldsymbol{x}_t \odot \hat{\boldsymbol{w}}_t \tag{2}$$

その後，特徴量 $\boldsymbol{y}_t$ を入力として区分化モジュールによる予測を行う．これにより，工程ごとに適切な特徴量を使用した予測が行われることが期待できる．

3.2　学習方法

損失関数を式 (3) と定義する．

$$L = L_{seg} + \lambda L_{ide} \tag{3}$$

ここで，λ は 2 つのモジュールのバランスを調節するハイパーパラメータである．選択モジュールは，各フレームの工程を識別する AS モデルを用いて構築する．よって，選択モジュールの損失関数 L_{ide} は式 (4) となる．ただし，L_{AS} は利用する AS モデルの損失関数を表す．

$$L_{ide} = L_{AS} \tag{4}$$

区分化モジュールは入力層に LassoLayer が接続されている．そのため，式 (5) によって区分化モジュールおよび LassoLayer の学習を行う．

$$L_{seg} = L_{AS} + \lambda_{L1}||\boldsymbol{W}||_1 \tag{5}$$

ここで，λ_{L1} は L1 正則化の強度を調節するハイパーパラメータである．LassoLayer の重みの L1 ノルム $||\boldsymbol{W}||_1$ を加えて学習を行うことで，重み $\boldsymbol{W}$ は重要度の低い入力がゼロとなるように縮小される．このようにして，LassoLayer は特徴量の選択を行う．ただし，損失関数に L1 項がある場合でも，学習によってパラメータを正確にゼロとすることは困難である．ここでは，[3] の方法を採用し，学習時にパラメータの絶対値がハイパーパラメータ λ_{L1} よりも小さい場合，その値を強制的にゼロとする．

3.3　先行研究の課題

先行研究 [13] は，フレームに対応する工程に応じて特徴量を選択することで予測に使用する特徴量を大幅に削減できることを示したが，より適切な特徴量を選択する観点では改善の余地がある．その理由として，学習時と推論時で同一の選択モジュールの予測結果をもとに重みフィルタの選択をしている点が挙げられる．すなわち，選択モジュールが誤った予測を行った場合，その後の重みフィルタの選択も誤るため，適切な特徴量を選択することができない．特に，学習時において選択モジュールの予測精度が不十分な場合，誤った工程の重みフィルタを選択して学習することになる．

4　提案手法

4.1　提案手法の概要

先行研究 [13] の課題を解決するために**重みフィルタの作成方法**，**学習方法**の 2 つの点の改善を行う．

まず，重みフィルタの作成方法について説明する．t 番目の入力フレーム $\boldsymbol{x}_t \in \mathbb{R}^N$ に対して，選択モジュールにより尤度 $\boldsymbol{z}_t$ の予測を行う．予測した尤度 $\boldsymbol{z}_t$ と LassoLayer の重み W から，入力フレーム $\boldsymbol{x}_t$ に適した特徴量を選択するための重みフィルタ $\hat{\boldsymbol{w}}_t$ を算出する．重みフィルタ $\hat{\boldsymbol{w}}_t$ は式 (6) のように定める．

$$\hat{\boldsymbol{w}}_t = \sum_{i=1}^{C} \alpha_{t,i} \boldsymbol{w}_i \tag{6}$$

また，$\alpha_{t,i}$ は式 (7) によって求められる．

$$\alpha_{t,i} = \begin{cases} 1 & (z_{t,i} \geq \sigma) \\ 0 & (z_{t,i} < \sigma) \end{cases} \tag{7}$$

ここで，σ は閾値を表す．上記の処理により，選択モジュールの予測尤度 $\boldsymbol{z}_t$ のうちから閾値 σ 以上の工程に対応する LassoLayer の重みを組み合わせることで，入力フレーム $\boldsymbol{x}_t$ に適した重みフィルタ $\hat{\boldsymbol{w}}_t$ を作成する．先行研究 [13] で使用されている予測工程に相当する重みフィルタ (式 (1)) ではなく，予測尤度をもとに複数の重みフィルタの和をとることで，選択モジュールが誤った予測を行った際の影響を緩和する狙いがある．

次に学習方法の改善について述べる．提案手法では，LassoLayer および 2 つのモジュールを全て同時に学習するのではなく，段階的に繰り返し学習する．これにより

モジュール間の相互の整合性を徐々に取ることが期待できる．表 1に示す 1 から 3 までのステップを損失が収束するまで繰り返し行うことで各モジュールを学習する．ここで，表 1中のモジュール名に対応する列は各ステップにおいてそのモジュールの学習を行うかどうかを表している．また，「重みフィルタの作成」は重みフィルタを作成する際に選択モジュールの出力尤度を使用するか，選択モジュールの出力尤度の代わりに正解ラベルを使用するかを表している．ステップ 1 では，正解ラベルを使用して重みフィルタを作成し，LassoLayer および区分化モジュールの学習を行う．正解ラベルを使用して LassoLayer を学習することで，工程ごとにより適切な特徴量を選択するための重みを学習させる．ステップ 2 では LassoLayer の重みを固定して区分化モジュールの学習を行う．重みフィルタの作成にはステップ 1 同様に正解ラベルを使用しており，区分化モジュールにおいて選択された特徴量を入力とした際の出力との対応を学習させる．ステップ 1 およびステップ 2 では，選択モジュールは LassoLayer に接続せずに元の AS モデルとして学習する．. ステップ 3 では選択モジュールおよび LassoLayer の重みを固定し，区分化モジュールのみを学習する．推論時と同様に選択モジュールの出力尤度を使用して重みフィルタを作成することで，正解ラベルを使用して学習した際の出力とのギャップを埋める目的がある．

5　実験設定と評価方法

5.1　使用する部品組立作業データ

先行研究 [13, 15, 16] において，工場での新人教育用の組立作業工程をベースとした部品組立作業を設計している．部品組立作業は，ガイドピンの取り付けやドライバーおよび電動ドライバーによるボルトの締め付け，ホースの取り付けなど，工場で作業をする上で必要となるさまざまな要素を含む 12 種類の行動から構成される．部品組立作業の内容を表 2，作業の例を図 2に示す．この作業を 20 代男女 4 名が合計 178 回実施し，図 3に示すように 3 台のカメラ (正面視座，右視座，左視座)，両手首および頭に装着した加速度・角速度センサー，右手首に装着したマイクロフォンによってデータを取得した．次に，タイムスタンプをキーとして取得し，それらを 1 つに結合した．これにより，マルチモーダルな時系列データの作成を行なった．

また，動画，加速度，角速度，音声の 4 種類の特徴量を使用した．カメラから取得したデータに対して OpenPose[17] を用いた処理を行い，作業者の各骨格点の 2 次元座標を取得した．取得した骨格点情報の例を図 4に示す．作業は主に上半身の動作によって行われるため首，右肩，右肘，右手首，左肩，左肘，左手首の 7 点の座標を抜粋して使用した．加速度および角速度は，左右手首及び頭のセンサーから取得したデータを使用した． それぞれのセンサーについて，加速度と角速度

図 2　部品組立作業の工程の例 [13]

ともに X 軸，Y 軸，Z 軸のデータを使用した．音声は

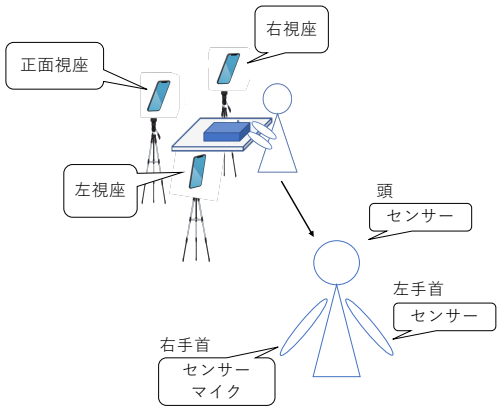

図 3　データの取得方法

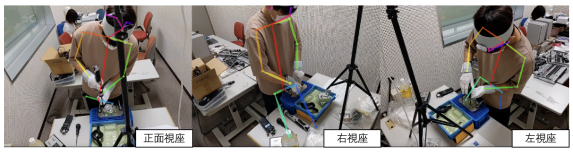

図 4　カメラから取得されるデータによって得られる骨格点情報

右手のマイクロフォンから取得したデータを 12 次元の MFCC 特徴量に変換して使用した．特に記述のない場合は動画，加速度，角速度，音声の 4 種類の合計 72 の特徴量を入力として使用している．

5.2　評価方法

本研究では，選択モジュールおよび区分化モジュールとして代表的な AS モデルである MS-TCN を使用する．また提案手法の有効性を以下の 3 点を通して評価する．

- 分節化モデルの精度比較
- 選択特徴量の可視化
- 選択特徴量の分布比較

分節化モデルの精度比較 まず，提案モデルの性能を精度の面で評価する．比較モデルとして元の MS-TCN，先行研究 [13] のモデルおよび最適モデルを使用する．最適モデルの結果は，学習時と予測時ともに選択モジュールの出力の代わりに正解ラベルを使用することによって得られる．つまり，選択モジュールが工程の予測を正確に行うことができた場合の結果とみなすことができる．これらの 4 つの手法を，各工程の F1 スコアの平均と各工程の予測に使用した特徴量の数の平均で比較する．

選択特徴量の可視化 学習後の LassoLayer の重みを可視化することで，工程ごとの特徴量の重要度の可視化を行う．各工程 (合計 12 工程) と入力 (3 台のカメラ，1 つの音声，3 つの加速度および角速度センサー) に対して LassoLayer によって選択された特徴量の割合を式 (8) によって算出する．

$$R_i^j = \frac{{Fsel}_i^j}{F_i^j} \tag{8}$$

ここで，工程番号 i，データソース j の特徴量の使用率を R_i^j，各入力の特徴量の合計 F_i^j のうち，選択された特徴量を ${Fsel}_i^j$ とする．求めた使用率が工程の特性によってどのように変化しているのか確認し，また，先行研究 [13] によって得られた使用率との比較を行う．

表 1　提案手法における各モジュールの学習のステップ

	選択モジュール	LassoLayer	区分化モジュール	重みフィルタの作成
1	学習	学習	学習	正解ラベル
2	学習	重み固定	学習	正解ラベル
3	重み固定	重み固定	学習	選択モジュールの出力尤度

表 2　部品組立作業の内容

工程番号	工程内容
0	ガイドピン取出・セット
1	オイルクーラー取出・O リングサブ取付
2	オイルクーラー取付
3	オイルクーラーボルト取出・挿入
4	オイルクーラーボルト仮締
5	オイルクーラーボルト締付
6	ガイドピン取出
7	オイルクーラーボルト取出・挿入
8	オイルクーラーボルト仮締
9	オイルクーラーボルト締付
10	ホース 1 取出・嵌込
11	ホース 2 取出・嵌込

選択特徴量の分布比較 特徴量選択の妥当性を定量的に評価するために選択された特徴量と選択されなかった特徴量の分布を比較する．適切な特徴量のみを選択することができていた場合，MS-TCN は選択した特徴量のみを使用して工程の予測ができる．すなわち，選択された特徴量はその工程を特徴づける十分な情報を持っていると言える．このことから，取得した骨格点に着目したとき，**選択された特徴量はその工程を特徴づける動きをしている**といった仮説が立てられる．図 5に仮説のイメージを示す．ある工程で選択された特徴量のデータ集

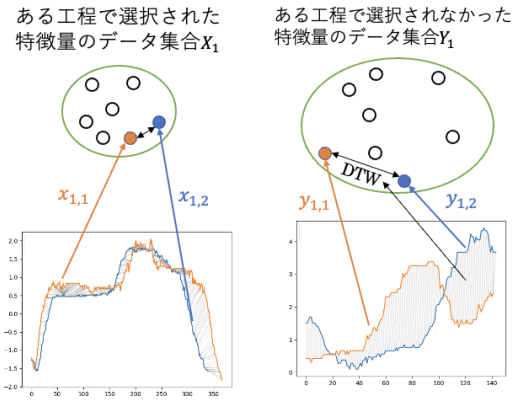

図 5　仮説のイメージ

合を $\boldsymbol{X}_i = \{x_{i,1}, ..., x_{i,50}\}$，ある工程で選択されなかった特徴量のデータ集合を $\boldsymbol{Y}_i = \{y_{i,1}, ..., y_{i,50}\}$ とするとき，$x_{i,j}(1 \leq j \leq 50)$ および $y_{i,j}(1 \leq j \leq 50)$ はそれぞれ対象工程の時系列データを表す．$x_{i,j}$ のそれぞれのデータ間の類似度を計算し，平均を求めることで $\boldsymbol{X}_i$ の分布の広がりを定量化する．**選択された特徴量はその工程を特徴づける動きをしている**場合，選択された特徴量の分布の広がりは選択されなかった特徴量の分布の広がりより小さいと考えられる．

実験は 4 人の作業者のデータ各 50 回分から，3 台の動画から取得されたデータの骨格点の特徴量のみを使用して行う．提案モデルを各作業者のデータを用いて学習する．また作業者ごとに，各工程において各骨格点の座標から x 軸と y 軸のそれぞれの動きの軌跡を正規化し，類似度を求める．類似度の計算には動的時間伸縮法 (*Dynamic Time Warping*，DTW)[18] を使用する．DTW とは時系列データ同士の距離や類似度を測る際に用いる手法である．2 つの時系列データの各点の距離を総当たりで求めて 2 つの距離が最短となるパスを見つけるため，長さや周期が違っても類似度を求めることができる．DTW は 2 つのデータ間の距離を求める手法であるため，各作業者で 50 回分のデータのうち 2 つを選択して距離を求める．50 回分のデータから 2 つのデータを選択する組み合わせは 1225 組あるが，そのうちから 100 組をランダムに選択して使用する．最後に，LassoLayer によって各工程で選択された特徴量と選択されなかった特徴量の DTW の値の平均を求め，比較を行う．

6　実験結果

6.1　分節化モデルの精度比較に関する実験結果

表 3　F1 スコアの平均

	F1 スコア	使用した特徴量の数
MS-TCN	0.956	72
先行研究 [13]	0.957	46.1
提案モデル	0.957	37.2
最適モデル	0.997	38.2

表 3は，各工程の F1 スコアの平均を示している．「使用した特徴量の数」は，各フレームの 72 の特徴量のうち，各工程の予測に使用される特徴量の数の平均を表す．提案手法の学習では，表 1の 1 から 3 までのステップを 3，4 回繰り返した合計のエポック数が先行研究 [13] のモデルの学習に必要なエポック数と同程度になるように各ステップのエポック数を設定した．これにより，1 から 3 までのステップが一周した時点で損失が収束することを防ぎ，3，4 回周した時点で学習が完了するように設定した．元の MS-TCN と比較すると，提案モデルは同程度の精度で予測を行なっているが，使用される特徴量は大幅に削減されている．また，先行研究 [13] と比較してもより少ない特徴量で予測を行うことができており，先行研究 [13] より解釈性の高いモデルであるといえる．

最適モデルは最も高い精度で予測を行なった．図 6は あるテストデータに対して，各モデルによって出力された尤度を比較したものである．元の MS-TCN および提案手法では工程の変わり目の部分で誤った予測が行われているのに対し，最適モデルでは工程の変わり目の部分でも正しく予測されており，この部分に大きな精度の差が出ていることが確認できる．本実験で使用しているデータセットは，予め定められたアノテーション規則に

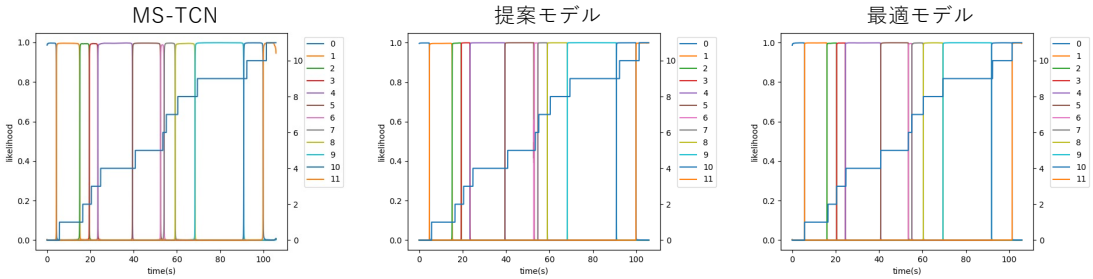

図 6　尤度グラフの比較

従って人手でアノテーションをおこなっている．そのため，アノテータの違いなどの要因によって工程の変わり目がある程度の範囲でずれる可能性が考えられる．このことから，元の MS-TCN や提案手法のように工程の変わり目において予測が実際の工程とづれてしまうことは当然の結果であると考えられる．一方で，最適モデルは選択モジュールの出力の代わりに正解ラベルを使用して特徴量選択をおこなってるため，予測のために間接的に正解ラベルの情報を与えられている．よって，工程の変わり目であっても正しく予測を行うことができていると推測される．以上のことから，提案モデルと最適モデルの間の精度の差は重大な問題ではないと考えられる．

6.2　選択特徴量の可視化に関する実験結果

表 4　各入力の特徴量の使用率 (動画および音声)

工程番号	動画 (正面視座)	動画 (右視座)	動画 (左視座)	音声
0	0.50 (0.57)	**0.59** (0.50)	0.52 (0.71)	**0.75** (0.72)
1	0.43 (0.60)	0.62 (0.71)	0.70 (0.74)	0.81 (0.81)
2	0.36 (0.57)	**0.71** (0.57)	**0.63** (0.55)	**0.63** (0.56)
3	0.57 (0.64)	**0.62** (0.60)	**0.66** (0.62)	0.65 (0.83)
4	0.64 (0.83)	0.79 (0.90)	0.66 (0.98)	0.75 (0.92)
5	0.48 (0.67)	0.63 (0.79)	0.62 (0.83)	0.48 (0.94)
6	0.25 (0.55)	0.41 (0.71)	0.48 (0.60)	0.31 (0.69)
7	0.50 (0.62)	**0.63** (0.62)	0.66 (0.69)	0.35 (0.86)
8	0.62 (0.57)	0.71 (0.76)	0.70 (0.79)	0.52 (0.75)
9	0.46 (0.52)	**0.66** (0.62)	0.57 (0.69)	0.60 (0.67)
10	0.46 (0.71)	0.54 (0.60)	0.59 (0.74)	0.46 (0.47)
11	0.32 (0.45)	**0.50** (0.48)	0.52 (0.67)	0.23 (0.39)

学習後の LassoLayer の重みを可視化することで，工程ごとに選択された特徴量の比率をデータソース毎に算出を行なった．表 4に動画および音声，表 5にセンサー (加速度，角速度) および音声の特徴量の使用率を示す．() の中の結果は，先行研究 [13] のものである．太字は先行研究 [13] の使用率要理も高いものを示している．

工程番号 0 は，ガイドピンを取り付ける工程である．この工程では，道具箱からガイドピンを取り出し，作業台に取り付けるという作業を行う．道具箱および作業台が設置されている位置が低いため，作業者の身体的特性により作業者の動きに大きなばらつきがある．このような作業者の身体的特徴に応じて動きが変化するような工程では，音声と比較して動画の使用率が低い傾向にあることが確認された．工程番号 10 および 11 はホースを取り付ける工程であり，他の工程で発生するような金属音が発生しない．音声がほとんどまたは全くないような工程では，音声に比べて動画の使用率が高くなる傾向があった．また，工程 1 は o リングを取り付ける作業であり，動きの種類が少なく，単純な動きが続く工程である．この工程では両手の加速度の使用率が高いことが確認できた．角速度はどの工程においても使用率が低い．よって，今回使用した 4 種類の特徴量のうち最も重要度が低いと考えられる．

先行研究の結果と比較した際に，提案手法は使用率がより少なく，使用している特徴量も厳選されていることがわかる．特に，正面視座の動画と角速度 (左手，右手，頭) の使用率が減少している．一方で，右視座の動画の使用率は先行研究と比較して増加している工程が多く，動画 3 視座の間での重要度をより強調した結果が得られたと考えられる．そこで提案手法の結果から重要度の低いと考えられる，正面視座の動画と角速度 (左手，右手，頭) の特徴量を除いた入力を使用して元の MS-TCN で予測を行い，全ての特徴量を使用した際の元の MS-TCN の結果と比較をおこなった．結果は表 6の通りである．正面視座と角速度以外の特徴量を入力として使用しても全特徴量を入力として使用した場合と同程度の精度で予測が行われていることが確認できた．以上より，提案モデルが示す特徴量の重要度は妥当であり，提案モデルによってデータ取得に使用するデバイスの選別を行うといった応用例の有効性を示した．

6.3　選択特徴量の分布比較に関する実験結果

結果を表 7に示す．表中の◯は選択された特徴量，×は選択されなかった特徴量を表す．

最適モデルは正解ラベルを与えられて重みフィルタの学習を行うため，最も適切な特徴量を選択することができるモデルであると考えられる．最適モデルの結果を見ると，どの作業者においても選択された特徴量の方が値が小さいことが確認できる．以上の結果から，仮説通りに「選択された特徴量はその工程を特徴づける動きをしている」と考えることができる．また，先行研究 [13] のモデルと提案モデルの結果を見ると，先行研究では選択された特徴量と選択されなかった特徴量の DTW の平均の大小関係が作業者によって異なるのに対し，提案モデルは最適モデル同様にどの作業者においても選択された特徴量の方が値が小さいことが確認できる．したがって，提案モデルは先行研究 [13] よりも適切な特徴量を選択できていると考えられる．

7　まとめと今後の課題

AS モデルに LassoLayer を組み込んだ先行研究 [13] が提案された．先行研究 [13] は特徴量利用の節減効果の面で課題があった．そこで，先行研究 [13] の課題を**重みフィルタの作成方法，学習方法**の 2 つの点で改善したモ

表 5　各入力の特徴量の使用率 (動画および音声)

工程番号	加速度 (左手)	加速度 (右手)	加速度 (頭)	角速度 (左手)	角速度 (右手)	角速度 (頭)
0	0.42 (0.67)	0.58 (0.78)	**0.58** (0.56)	**0.25** (0.22)	0.42 (0.56)	**0.58** (0.33)
1	0.83 (1.00)	**0.83** (0.78)	1.00 (1.00)	0.50 (1.00)	0.42 (0.78)	**0.42** (0.33)
2	**0.75** (0.67)	0.50 (0.56)	**0.67** (0.44)	0.25 (0.44)	0.17 (0.33)	0.25 (0.56)
3	**0.67** (0.44)	0.58 (0.78)	**0.58** (0.44)	**0.50** (0.44)	0.33 (0.56)	0.58 (0.78)
4	0.75 (0.89)	0.75 (1.00)	0.75 (1.00)	0.08 (0.56)	0.17 (0.67)	0.33 (0.67)
5	0.42 (0.78)	0.33 (0.56)	0.25 (0.67)	0.17 (0.33)	0.25 (0.78)	0.42 (0.56)
6	0.17 (0.44)	0.25 (0.67)	0.33 (0.44)	0.08 (0.44)	0.17 (0.22)	0.42 (0.56)
7	**0.50** (0.44)	0.67 (0.67)	0.42 (0.78)	0.17 (0.78)	0.42 (0.56)	0.25 (0.44)
8	0.67 (0.78)	0.58 (0.67)	0.42 (0.67)	0.17 (0.33)	0.00 (0.56)	0.25 (0.56)
9	0.50 (0.56)	0.33 (0.78)	0.50 (0.56)	0.17 (0.44)	0.08 (0.11)	**0.17** (0.11)
10	**0.50** (0.33)	0.42 (0.67)	**0.42** (0.11)	**0.33** (0.22)	0.08 (0.33)	0.17 (0.33)
11	0.08 (0.11)	0.25 (0.67)	**0.42** (0.33)	0.00 (0.22)	0.00 (0.22)	0.08 (0.11)

表 6　入力に全特徴量を使用した場合と正面視座と角速度以外の特徴量した場合の精度比較

使用した特徴量	F1 score
全特徴量	0.956
正面視座と角速度以外の特徴量	0.961

表 7　DTW の平均の比較

		先行研究 [13]	提案モデル	最適モデル
作業者 A	○	2.52	1.98	2.41
	×	1.80	3.20	2.70
作業者 B	○	4.05	4.16	3.41
	×	5.89	4.19	5.34
作業者 C	○	3.32	2.80	2.95
	×	2.43	6.98	4.13
作業者 D	○	1.50	1.29	1.39
	×	1.49	2.36	1.68

デルの提案を行った．精度の面では，提案モデルでは元の MS-TCN，先行研究 [13] のモデルと比較して，より少ない特徴量を使用して同程度の精度で予測を行うことが確認できた．また，各工程に対応するフィルタの重みを可視化することで，工程ごとに選択された特徴量の比率をデータソース毎に算出し，実際の製造現場での作業を想定したデータセットを使用した実験によって，モデルの妥当性を定性的に評価した．結果として，工程の特徴に合わせて使用される特徴量に変化があることが確認できた．選択された特徴量と選択されなかった特徴量の類似度を比較することで，**選択された特徴量はその工程を特徴づける動きをしている**という仮説を立証し，先行研究 [13] より適切な特徴量を選択していることを示した．

今後の課題として，選択モジュールと区分化モジュールに使用する AS モデルの検討を行う予定である．本稿で使用した MS-TCN 以外のモデルを使用して精度の比較を行い，それぞれのモジュールに使用するべき最適な AS モデルの組み合わせについて検討する．また，各作業者の動きの違いの分析のための提案手法の応用方法についての検討を行う予定である．作業者ごとに提案モデルを学習した際に選択される特徴量の違いの分析や，作業者の作業のスコアリングのためのシステムへの応用が可能か検証する．

参考文献

[1] 西村雅史他: ものづくり現場における組立作業の 行動認識, 人工知能学会学会誌特集号, Vol.5, 2022.

[2] 八田俊之他: 生産現場の効率化に貢献する作業分 析システム, 三菱電機技報, 2021.

[3] A. Sudo, et al.: LassoLayer: Nonlinear Feature Selection by Switching One-to-one Links, arXiv preprint arXiv:2108.12165, 2021.

[4] C. Lea, et al.: Temporal convolutional networks for action segmentation and detection, CVPR, 2016.

[5] Y. A. Farha, et al.: MS-TCN: multi-stage temporal convolutional network for action segmentation, *CVPR*, 2019.

[6] S. Li, et al.: MS-TCN++: multi-stage temporal convolutional network for action segmentation, TPAMI, 2020.

[7] Yi, F., et al.: Asformer: Transformer for action segmentation. arXiv preprint arXiv:2110.08568, 2021.

[8] Behrmann, N., et al.: Unified fully and timestamp supervised temporal action segmentation via sequence to sequence translation, In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXV (pp. 52-68). Cham: Springer Nature Switzerland.

[9] A. Vaswani, et al.: Attention is all you need, Advances in neural information processing systems 30, 2017.

[10] Y. Ishikawa, et al.: Alleviating over-segmentation errors by detecting action boundaries, WACV, 2021.

[11] G. Chandrashekar, et al.: A survey on feature selection methods, Computers & Electrical Engineering, 40(1), 16-28, 2014.

[12] R. Tibshirani: Regression shrinkage and selection via the LASSO, Journal of the Royal Statistical Society, Series B (Methodological), pages 267–288, 1996.

[13] 清水裕斗他: 特徴量選択を伴う時間畳み込みネットワークを用いた組立作業工程の行動分析，第 37 回人工知能学会全国大会，2023.

[14] 久保莞太他: 自動車組立作業における時系列行動セグメンテーション手法の比較検討，第 37 回人工知能学会全国大会，2023.

[15] 武井久実他: 慣性情報と音情報を用いた作業行動の自動分節化，第 84 回情報処理学会全国大会，2022.

[16] 中村圭佑他: 作業分節化のための Attention 付き TCN の検討，第 126 回知識ベースシステム研究会, 2022.

[17] Z. Cao, et al.: OpenPose: realtime multi-person 2D pose estimation using part affinity fields, Pattern Analysis and Machine Intelligence, 2019.

[18] F. Petitjean, et al.: A global averaging method for dynamic time warping, with applications to clustering, Pattern recognition 44.3, 2011.