Vol.2021-CVIM-224 No.2
2021/1/21

発表概要

異なる粒度の作業分析のための動作系列Sequence to Sequenceモデル

惠本 序珠亜[1,a)]　西行 健太[1,b)]　戎野 聡一[1,c)]　木下 航一[1,d)]

工場の生産性向上には作業の分析が必要であるが，作業手順のミス検知，効率的な手順の探索など，作業分析の目的や対象により，認識したい作業の粒度は異なる．しかし，全ての粒度に対応するため，各粒度の動作をラベル付けした学習データを用意することは難しい．そこで，ある粒度の動作を認識する識別器を作成した後，少数データを用いて，それらの識別器の出力を粒度の粗い作業に変換する手法を提案する．その一例として，「把持」「運搬」「調整」などの粒度の細かい動作系列を「箱を台に固定する」「製品を箱に入れる」などの粒度の粗い作業系列に変換するモデルを開発した．提案手法は，42 名の作業者から撮影した工場での作業を模した 5 つの工程の動画データで評価した．1 名の作業者の動画データだけを用いて学習した結果，5 工程中 4 工程において 90 %を超える精度を達成した．

Presentation Abstract

Joshua Emoto[1,a)]　Kenta Nishiyuki[1,b)]　Soichi Ebisuno[1,c)]　Koichi Kinoshita[1,d)]

Work analysis is essential method to improve the productivity of manufacturing. In work analysis, the granularity of the work to be analyzed varies depending on the purpose and objective of work analysis, such as detecting irregular procedures and designing efficient procedures. However, it is difficult to prepare the data of each granularity of the work for machine learning. In this paper, we propose a method to convert the fine-grained motion sequence into coarse-grained motion sequence using a small amount of data. We present the sequence to sequence model that converts fine-grained sequence such as ”grasping”, ”transporting”, and ”adjusting” into coarse-grained work series such as ”fixing the box to the jig” and ”put the product in the box”. The model was evaluated on simulating data of five work procedure in a factory taken from 42 workers. The proposed method achieved 90% accuracy in 4 out of 5 work procedure using only single video data from one worker.

1. はじめに

製造現場では生産性向上のため，Industrial Engineering (IE) に基づく作業分析が行われている [3]．IE に基づく作業分析は時間と工数がかかるが，現代の製造現場は常に状態が変化することから，改善のサイクルを早く回すことが求められている [2]．そのため，作業分析を自動化する技術が必要とされている．

作業分析による改善の効果を定量的に評価するためには，作業時間を計測する必要がある [3]．動作認識技術の発展により，作業中の動画（作業動画）を解析することで，作業時間を自動的に測定する作業認識手法が提案されている [7], [8]．

作業分析では目的によって，考慮する作業の粒度が異なる．例えば，作業者の習熟度を評価したり，効率的な動き方を分析する場合には，「把持」「運搬」「調整」のような粒度の細かい動作を分析する．効率的な作業手順の探索や，作業手順の欠陥を分析する場合には，「箱を台に固定する」「製品を箱に入れる」などの粒度の粗い作業を分析する．従

[1] オムロン株式会社
Omron, Kidugawadai, Kyoto 619–0225, Japan
[a)] joshua.emoto@omron.com
[b)] kenta.nishiyuki@omron.com
[c)] soichi.ebisuno@omron.com
[d)] koichi.kinoshita@omron.com

来の作業時間の自動測定技術 [7], [8] では，異なる作業粒度に対応していないため，作業粒度を考慮した分析のコストが高い．

本論文では，異なる粒度の作業分析に対応するために，少量のアノテーションだけで粒度の細かい動作系列を粒度の荒い作業系列に変換する Seqence to Sequence (Seq2Seq) モデルを提案する．提案手法は 5 工程の作業を模擬したデータセットを用いて評価する．

2. 従来手法

2.1 系列変換

提案手法の Seq2Seq モデルは，時刻順に並んだ動作系列を作業系列に変換する．系列変換に有効なモデルとして，1 次元畳み込みニューラルネットワーク (1DCNN)[5] や Long short-term memory (LSTM)[4]，Transformer[6] が提案されている．

1DCNN は畳み込みにより系列の時間的な特徴を抽出する．出力系列には，畳み込み層のカーネルサイズに依存して，出力の各時刻近傍に対応する入力系列の情報が反映される．LSTM は再帰的な構造により時間的な特徴を抽出する．出力系列には各時刻より過去の情報が反映される．Transformer は Self-attention により時間的な特徴を抽出する．出力系列の各時刻には，入力系列全体の情報が反映される．

2.2 作業認識

工場作業者の作業分析を自動化するため，動作認識技術を応用した作業認識手法が提案されている．

清水ら [8] は，動画から抽出した骨格情報に基づいて，あらかじめ定義された作業内容を推定する手法を提案している．清水らの手法では，骨格情報と作業手順の組み合わせから成る作業動作モデルを作業者ごとに作成する．また，異なる工程や粒度の作業を認識するには，それぞれに学習が必要である．清水らの手法は，作業者と工程，作業粒度に依存しており，異なる作業粒度の認識には高いアノテーションコストが必要となる．

内田ら [7] は，把持，運搬など工程間で共通する動作を認識する手法を提案している．粒度の細かい動作を認識対象とすることで工程への依存を軽減したが，異なる粒度の作業分析には対応できない．

3. 提案手法

本論文では，文献 [7] における把持や運搬などの作業の共通動作の系列（粒度の細かい動作系列）からより抽象的な作業の系列（粒度の粗い作業系列）に変換する Sequence to Sequence(Seq2Seq) モデルを提案する．

提案手法は，内田らの手法 [7] の認識対象である「把持」「運搬」「調整」のような抽象的な情報の系列を入力とする．

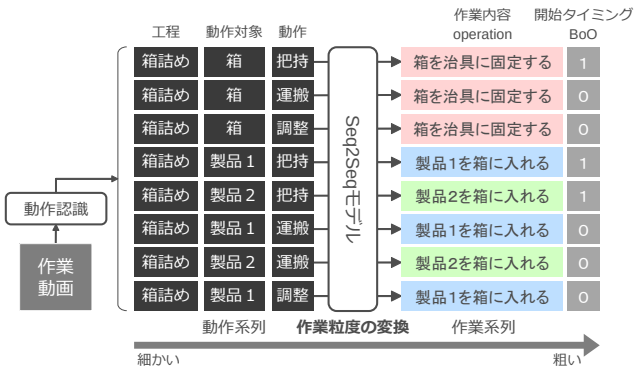

図 1 作業粒度の変換と入出力データ

これにより，骨格情報のような工程や作業者に依存した情報を入力とした場合に比べ，Seq2Seq モデルは，少量のアノテーションによる学習が期待できる．提案手法では，アノテーションの負荷を抑えつつ，異なる作業粒度に対応する．

3.1 入出力データ

図 1 に提案手法による作業粒度変換の流れと入出力データを示す．入力データは，作業動画中に行われた動作の種類，その対象となった物体や作業工程の種類などのメタ情報が，時系列順に並んだ系列データである．これらの情報は自然言語処理における単語と同様，one-hot ベクトルで表現される．出力データは，入力の各動作１つ１つに対応する作業の種類と，その動作が作業の開始タイミングであるかを表す 0/1 の値である．入力の系列データは人手によるアノテーションにより用意する．

3.2 ネットワーク構造

図 2 に Seq2Seq モデルのネットワーク構造を示す．Seq2Seq モデルは，入力として，動作系列（把持，運搬，調整などの共通動作の系列）と動作対象の製品の種類や工程の種類など，メタ情報の系列を受け取る．これらの入力系列を共通の埋め込み空間に埋め込み，算出した埋め込み表現を各時刻ごとに足し合わせる．足しあわされた各時刻ごとの埋め込み表現は，埋め込み表現の系列として系列特徴抽出器に入力される．例えば，時刻１で「箱」を「把持」し，時刻２で「箱」を「運搬」した場合，時刻１の「箱」と「把持」の埋め込み表現を足し合わせた値と，時刻２の「箱」と「運搬」の埋め込み表現を足し合わせた値の系列が系列特徴抽出器に入力される．入力次元数は時刻 × 埋め込み次元数であり，時刻は可変である．

入力系列のメタ情報の埋め込み表現は，系列特徴抽出器に入力される前に全て足しあわされるため，作業の推定に役立つと考えられる情報（作業工程の種類や動作の対象となった物体の種類，動作を行った手が右手か左手か等）を自由に設定できる．

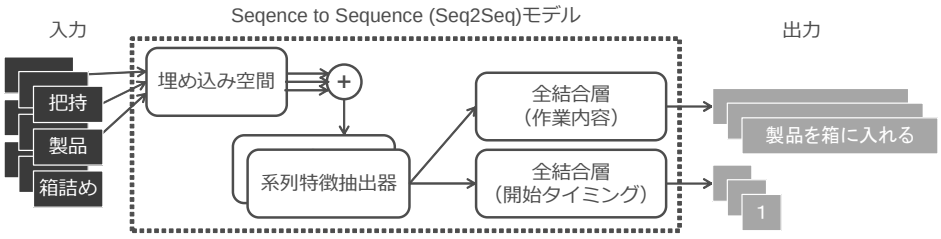

図 2 提案手法のネットワーク構造

系列特徴抽出器には，入力系列と同じ長さの系列を出力する任意のネットワーク構造を用いる．提案 Seq2Seq モデルにおいて，系列特徴抽出器以外の処理は時刻に独立した処理である．系列特徴抽出器は入力系列の時間的特徴を抽出する役割を担っている．2.1 節で述べたように，ネットワーク構造の種類によって，抽出される時間的特徴が異なる．

系列特徴抽出器よって抽出された系列特徴は 2 つの全結合層に入力される．全結合層は，系列特徴の各時刻ごとで独立に適用され，各時刻ごとの作業内容と開始タイミングに変換される．本ネットワーク構造は可変長の入力系列に対応している．

4. 実験

本論文では，Seq2Seq モデルの入力を，動作，動作対象の物体，工程の種類の 3 つとする．動作が 7 種類，物体が 11 種類，工程が 5 種類の合計 23 種類を共通の埋め込み空間に埋め込む．出力する作業の種類は 5 工程合わせて 19 種類である．動画に対して人手で付与した動作と作業，メタ情報のアノテーションを用いて，Seq2Seq モデルを学習させ，その精度を評価する．

4.1 実験設定

本論文では系列特徴抽出器として，1DCNN と LSTM，Transformer の 3 つを用いる．

系列特徴抽出器のネットワーク構造は，それぞれのパラメータ数が同じになるように調整する．1DCNN の層数は 2，フィルタ数は 9，カーネルサイズは 3 とする．LSTM の層数は 2，全結合層のユニット数は 5 とする．Transformer の層数は 2，Attention ヘッド数は 2，全結合層のユニット数は 16 とする．1DCNN，LSTM，Transformer のパラメータ数は，それぞれ 691，690，695 となる．

1DCNN については，各畳み込み層でゼロ詰めすることで入力と出力の系列数を同数とする．Transformer については，系列特徴抽出器に Transformer のエンコーダ層のみを用いる．また，埋め込み空間で算出した値には Positional Encoding を加える．

埋め込み空間の次元数は 4 次元とする．活性化関数には Rectified Linear Unit，損失関数には Cross Entropy Loss を用いる．

表 1 データセットに含まれるデータの数

	工程	作業者	単位作業	動作	作業
学習	ケース嵌合	14	78	1788	546
	箱詰め	14	160	2735	800
	ネジ締め	13	34	1971	805
	ラベル貼り	13	52	1126	416
	ワイヤーハーネス	7	42	1576	252
試験	ケース嵌合	15	84	1925	586
	箱詰め	15	183	3152	915
	ネジ締め	13	39	2172	896
	ラベル貼り	15	51	1106	408
	ワイヤーハーネス	6	36	1327	216
合計		-	759	18878	5840

4.2 データセット

工場の作業を模擬した 5 つの工程の動画を撮影し，データセットを作成した．撮影した工程は，ケース嵌合，箱詰め，ネジ締め，ラベル貼り，ワイヤーハーネスの 5 つである．1 種類の工程において，1 製品を組み立てる作業の系列を単位作業とする．例えば，ケース嵌合工程であれば，部品置き場から部品を取り，部品を組み合わせ，ケースを組み立て，完成品置き場に完成品を格納するまでが 1 つの単位作業である．単位作業中の部品を取ることやケースを組み立てることが 1 つの作業である．作業で部品を取る中で，部品を把持することや運搬することが 1 つの動作である．

表 1 に各工程で撮影した作業者と単位作業，動作，作業の数を示す．学習データは 21 名，試験データは 21 名の計 42 名の作業者が複数の工程で作業したデータを収録した．学習データと試験データにはそれぞれ異なる作業者のデータが含まれる．後述する学習では，表 1 の学習データを，ネットワークの学習用と検証用に分ける．

4.3 学習

学習データの作業者のうち，各工程の作業者 1 名のデータのみを学習データとし，残りの作業者のデータを検証用とする．学習に用いる作業者はランダムに選択する．その結果，学習用の単位作業数はケース嵌合が 5 回，箱詰めが 9 回，ネジ締めが 3 回，ラベル貼りが 5 回，ワイヤーハーネ

表 2 系列特徴抽出器別の精度

		精度 [%]				
		ケース嵌合	箱詰め	ネジ締め	ラベル貼り	ワイヤーハーネス
1DCNN	学習	100	100	100	100	100
	試験	93.81	97.68	76.60	98.09	93.73
LSTM	学習	31.96	76.47	74.82	93.59	47.04
	試験	31.68	76.39	59.93	93.99	50.09
Transformer	学習	100	94.12	79.69	100	89.55
	試験	94.37	93.97	66.89	99.91	93.63

スが 6 回となった．作業時間としては，ワイヤーハーネス工程は 1 回の単位作業の時間が長く約 13 分，その他の工程は，およそ 2 分から 2 分半ほどのデータとなった．3 種の系列特徴抽出器の学習には，同一の学習用と検証用データの組み合わせを用いる．

最適化手法には確率的最急降下法（SGD）を用いる．学習率は 4，最大エポック数は 1000 とする．学習率は 10 エポックごとに 0.95 を係数として減衰させる．ドロップアウト率は 0.2 とする．10 エポックごとに検証用データで損失を計算し，損失が最小となった時点のモデルを用いて，試験データにより評価する．

4.4 結果

Seq2Seq モデルによる作業の推定精度に基づいて，提案手法の性能を評価する．表 2 に学習データと試験データにおける精度を示す．精度は単位作業ごとの正解率の平均として式 1 を用いて算出する．N は単位作業数，i は各単位作業の番号を表す．

$$\text{精度} = \frac{1}{N}\Sigma_i^N \frac{\text{正解数}_i}{\text{動作数}_i} \tag{1}$$

表 2 の試験データにおける精度を見ると，1DCNN と Transformer はネジ締め工程以外で，90 %以上の精度で作業を推定できることがわかる．一方，LSTM は 1DCNN と Transformer と比べて精度が低いとわかる．

1DCNN については，学習データにおける精度が全工程で 100 %に達していることから，過学習していると考えられる．

4.5 考察

本論文で用いた 1DCNN は，2 層の畳み込み層でカーネルサイズが 3 のため，出力系列には各時刻の近傍 5 点の入力系列の情報が反映される．試験データにおける精度が高かったことから，作業系列の推定には出力の近傍の動作系列の情報で十分であると考えられる．

LSTM は試験と学習データの両方で全体的に低い精度である．3 つの系列特徴抽出器は，学習の最大エポック数が同じで，パラメータ数もほぼ同数であるが，LSTM は 1DCNN や Transformer と異なり，出力系列には過去の時刻の情報のみが反映されていることから，作業系列の推定には未来の時刻の情報が重要であると考えられる．

全工程で 100%の精度となり過学習した 1DCNN と比べて，Transformer の学習データにおける精度は，試験データの精度と同様の傾向であり，過学習が発生していないと考えられる．Transformer は出力の各時刻に入力系列全体の情報が反映されることから，より汎化性能の高いモデルが獲得されたと考えられる．

どの Seq2Seq モデルでも試験データにおいて精度が低かったネジ締め工程については，提案手法のモデルでは推定困難なデータが含まれていたと考えられる．アノテーションの見直しやモデル構造の変更による表現力の向上，動作・物体・工程情報以外のメタ情報（画像特徴量など）を入力として利用するなどの工夫が必要と考えられる．

5. おわりに

動作系列から作業系列を推定する Seq2Seq モデルを提案し，独自に作成したデータセットで，その性能を評価した．その結果，作業者一人分のデータの学習で，5 工程中 4 工程において 90%以上の精度で動作系列から作業系列を推定できることを確認できた．

本論文では Seq2Seq モデルの学習の際に，入力データとして人手で作成した動作や動作対象物体のアノテーションを利用した．今後は動作認識や物体認識モデルと Seq2Seq モデルを組み合わせることで，作業動画から直接異なる粒度の作業分析を可能にする作業系列推定モデルを開発する．

参考文献

[1] Yasuyo Kotake, Danni Wang, Hiroshi Nakajima: Evaluating Worker's Proficiency from Body and Eye Movements in Manufacturing Operations, IEEE International Conference on Systems, Man, and Cybernetics(2018).
[2] 香川博昭：実践 IE の進め方，日科技連 (2007).
[3] 藤井春雄：よくわかる「IE 七つ道具」の本，日刊工業新聞社 (2011).
[4] Ilya Sutskever, Oriol Vinyals, Quoc V. Le: Sequence to Sequence Learning with Neural Networks, 27th International Conference on Neural Information Processing Systems(2014).
[5] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, Yann N. Dauphin: Convolutional Sequence to Sequence Learning, 34th International Conference on Machine Learning(2017).
[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob

Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser: Attention Is All You Need, 31st International Conference on Neural Information Processing Systems(2017).
[7] 内田 滋穂里, 西行 健太, 和田 洋貴, 木下 航一：工程依存性を考慮した作業者動作識別モデル, 画像の認識・理解シンポジウム MIRU(2020).
[8] 清水 尚吾, 草野 勝大, 奥村 誠司, 小林 拓椰, 青木 義満：骨格情報を用いた作業動作分析手法, 画像センシングシンポジウム SSII(2020).