特集記事

AIを活用した多変量時系列データの解析技術と状態監視への応用

地方独立行政法人東京都立産業技術研究センター、
東京大学
三木　大輔　Daisuke MIKI

1．はじめに

近年、監視カメラの設置が進められており、蓄積される膨大な映像の中から異常や人物の不審動作を自動的に検知する技術が求められている。最近ではAIを活用することで映像中の人物の動作を理解する取り組みが多く報告されており、このような映像解析技術を作業安全監視に応用できれば、保全作業における事故の発生を低減できる可能性がある。具体的には、人物の危険な動作や通常動作からの逸脱を検出・評価し、労働環境を改善する取り組みが考えられる。人物の動作データのような多変量時系列データからの異常検知は、状態監視や診断を実現するために重要な技術であり、統計的手法や機械学習手法等に基づくさまざまな手法が提案されている。特に深層ニューラルネットワーク（Deep Neural Network, DNN）を用いた手法は、その表現能力の高さから多変量時系列データを解析するうえで有用な手法として注目されている。最近ではこのようなDNNに基づく解析手法全般をAIと呼ぶことも多い。そこで本稿では映像から人物動作に関する特徴量を抽出するDNNについて述べた後、多変量時系列データから異常を検知可能なDNNおよびその学習手法を紹介する。

2．映像解析のためのDNN

建設作業現場において安全衛生管理体制の確立が求められているが、厚生労働省の報告では，2018年度の労働災害による死亡者数は約900人に上る[1]。このような事故を引き起こす要因の一つに労働環境に潜む危険、たとえば、滑りやすい床や炎天下・高温下などの過酷な労働環境があり、これらを原因とした滑り、つまずき、踏み外しなどの小さな異変が積み重なることで、転倒や機器の操作ミスを引き起こし、最悪の場合事故に至る。このような事故を未然に防ぐには、作業中の人物の危険な動作や，通常動作からの逸脱を迅速に検知し，労働環境を改善することが有効と考えられる。例えば監視カメラによって日々収集される膨大な映像から異常を自動で検知することができれば、安全対策をより強固にできる可能性がある。Chenらは深層学習を応用した映像解析によって作業者が保護具等を適切に着用し、安全な状態にあるか検知する技術を提案している[2]。このような映像解析技術を応用することで人物の危険な動作等を検知できれば労働環境の改善につながる可能性がある。

図１　保全作業における労働安全

さらに、原子力関連施設における映像解析技術の活用方法として核セキュリティへの応用が考えられる。妨害破壊行為を企てる人物から施設を守る上で監視カメラの利用が有効と考えられるが、妨害破壊行為の検知には早期の対応が必要であり、監視すべき映像が膨大であるため、映像を自動的に解析する技術が求められる。上記のような映像解析技術を応用することができれば、人物の通常動作からの逸脱の程度（重大さ）を定量化することで、異常を検知できる可能性がある。

そこで筆者らは、AIを活用することで映像から人物動作の特徴を抽出する手法を確立した後、その特徴から人物動作の異常を検知するための多変量時系列データ解析技術の開発を行った。

2.1 人物動作特徴量の抽出方法

映像は色情報をもつ画素が高さH、幅Wの格子状に

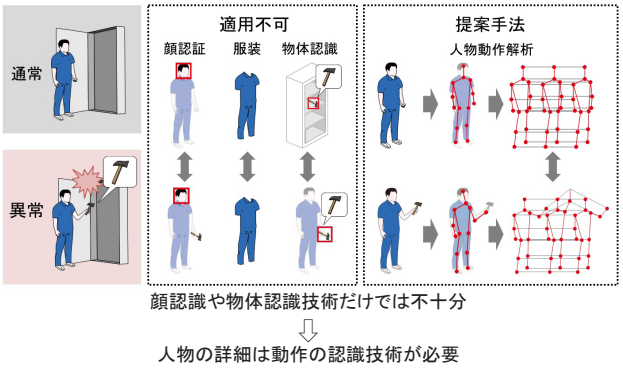

図２ 映像解析技術を活用した異常検知

配列された画像 $\mathbf{I} \in \mathbb{R}^{W \times H}$ の時系列データとして扱うことができるが、一般に画像のもつ情報は高次元であるため、事前にその特徴を適切に抽出し、次元削減を行う必要がある。そこで映像中の人物の動作はその人物の持つ関節位置座標集合の時系列情報、すなわち姿勢の時間変化として表現できることに着目した。そこで、人物動作を表す特徴量として２次元 RGB 映像中の人物の関節位置座標を推定した後、２次元的な関節位置座標と、事前に学習された３次元人物姿勢モデルから３次元的な人物姿勢情報および実空間上の人物の位置を推定することで、人物動作特徴量を抽出する方法を採用した。

映像中の人物の関節位置の推定には、畳み込みニューラルネットワーク（Convolutional Neural Network, CNN）に RGB 画像を入力し、画像の畳み込み、プーリング、および活性化を繰り返すことで、出力層では人物の関節が存在する確信度を示す高さ w、幅 h（$w \le W, h \le H$）の確信度マップを推定する構成とした。CNN の学習のため、人物を含む画像およびその関節位置座標に関するアノテーションが対となったデータセット [3] を利用した。データセットに含まれるアノテーションをもとに、j 番目の関節位置を示す確信度マップ $\mathbf{S}_j^*(\mathbf{p}) \in \mathbb{R}^{w \times h}$ は真の関節位置座標 $\mathbf{y}_j^*$ および確信度マップ中における各画素の座標 $\mathbf{p}$ を用いて

$$\mathbf{S}_j^*(\mathbf{p}) = \exp\left(-\frac{\|\mathbf{p}-\mathbf{y}_j^*\|_2^2}{\sigma}\right) \tag{1}$$

で表される。本研究では、CNN に RGB 画像を入力した際の推定値 $\mathbf{S}_j(\mathbf{p})$ と真値 $\mathbf{S}_j^*(\mathbf{p})$ の間における損失関数

$$\mathcal{L}_{pose} = \Sigma_j \|\mathbf{S}_j(\mathbf{p}) - \mathbf{S}_j^*(\mathbf{p})\|_2^2 \tag{2}$$

を最小化するように CNN のパラメータを最適化した。

以上により得られた２次元の間接位置情報を用いた人物の３次元姿勢の推定では、あらかじめ用意されたの３次元人物姿勢モデルに対して t 分布型確率的近傍埋め込み法による次元削減を行った後、EM アルゴリズムによるクラスタリングを行い、平均姿勢 $\mathbf{Y}_n^*$ および正規直交行列 $\mathbf{e}_n$ を求めた。次に、平均姿勢 $\mathbf{Y}_n^*$ の２次元平面への射影 $\mathbf{y}_n^*$ および先の注目点検出器によって推定された人物の画像中での２次元姿勢 $\mathbf{y}$ を用い

$$\arg\min_{n,\mathbf{a},\mathbf{R}} \|\mathbf{y} - \mathbf{y}_n^*(\mathbf{a}, \mathbf{R})\|_2^2 \tag{3}$$

を満たす平均姿勢のインデックス n、正規直交行列 $\mathbf{a}$ および回転行列 $\mathbf{R}$ を求め、対応する３次元での人物姿勢

$$\mathbf{Y}_n(\mathbf{a}, \mathbf{R}) = \mathbf{R}(\mathbf{Y}_n^* + \mathbf{a} \cdot \mathbf{e}_n) \tag{4}$$

を得た。ここで、平均姿勢 $\mathbf{Y}_n^*$ の２次元平面への射影 $\mathbf{y}_n^*$ は投影行列 $\mathbf{\Pi}$ を用いて

$$\mathbf{y}_n^*(\mathbf{a}, \mathbf{R}) = \mathbf{R}\mathbf{\Pi}(\mathbf{Y}_n^* + \mathbf{a} \cdot \mathbf{e}_n) \tag{5}$$

で表される。

2.2 人物動作特徴量の抽出実験

以上で提案した人物動作特徴量の抽出に関する評価として 3D コンピュータグラフィックスを用いて人物が歩行する映像を生成し、人物の姿勢および位置に関する真値と推定値の比較を行った。モーションキャプチャデータとして CMU Mocap データセット [4] を用い、データセットから、自然な歩行動作（Subject 91, trial 2）および不自然な歩行動作（Subject 91, trial 18）から人物動作の特徴量（体の向き、視線方向、人物の移動速度）の抽出を行い真値と比較した。

抽出された特徴量の真値との比較結果を図 3 に示す。それぞれの動作に関する特徴を可視化することができ、特に不自然な歩行動作において体の向きや視線方向、移動速度の変動に対して特徴的な傾向を把握することが可

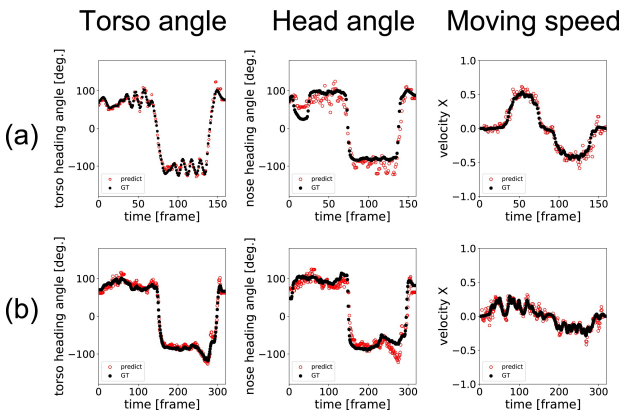

図３ 不自然な動作を (a) 含まないおよび (b) 含む
人物動作データから抽出された特徴量

能であった。次節では抽出された特徴量の解析手法について述べる。

３．多変量時系列解析のための DNN

時系列解析に用いられる一般的な DNN の学習では、一連の時系列データに対し、その特性を決定付ける上で重要な箇所に対してアノテーションを付与し学習に供することが有効であるが、その作業は煩雑かつ困難なものである。このような課題を解決するため、筆者らはマルチインスタンス学習に着想を得たランク学習 [5] を参考にすることで DNN の学習を行うための新たな損失関数を提案した。

3.1 異常検知のための DNN 学習方法

提案する DNN は長さ T の時系列データ $X=\{x^1,\cdots,x^T\}$ を入力した際に $y=\{y^1,\cdots,y^T\}$ を出力する構造とし、入力データにおける異常の有無は既知であるが、それらが含まれる箇所は未知であるものとする。DNN の学習では、まず時系列データに異常を含む(正)または、含まない(負)によってデータを各クラスに分割し、それぞれのデータを DNN に入力した際に、正のデータに対して高い値、負のデータに対して低い値が出力されることを目標に DNN のパラメータを最適化する。各データにおいてその特性を決定付ける上で重要な箇所は未知であるため、ここでは各データにおけるそれぞれ最大の値に着目し、

$$\max_{t\in\{1,\dots,T\}} y_{pos}^t > \max_{t\in\{1,\dots,T\}} y_{neg}^t \tag{6}$$

を満たすように DNN を学習する。ここで、y_{pos}^t および y_{neg}^t は正および負のそれぞれのデータを DNN に入力した際に出力される $t\in\{1,\cdots,T\}$ 時点における推定値である。このような条件を満たす DNN の学習では、損失関数

$$\mathcal{L} = \max\left(0, 1 - \max_{t\in\{1,\dots,T\}} y_{pos}^t + \max_{t\in\{1,\dots,T\}} y_{neg}^t\right) + \lambda. \tag{7}$$

を最小化するように DNN のパラメータを最適化する。ここで λ は

$$\lambda = p_1 \sum_{t=1}^{T-1}\left(y_{pos}^t - y_{pos}^{t+1}\right)^2 + p_2 \sum_{t=1}^{T} y_{pos}^t. \tag{8}$$

で表される正則化項であり、学習で用いるデータへの過学習を防ぎ、学習を安定させるために導入した。2 つの項はそれぞれ、平滑化項、およびスパース化項である。平滑化項は時間方向の出力値の変動を低減し、スパース化項は全体的な出力値を小さくすることで誤検出を低減する効果がある。提案する DNN の学習は (7) 式の損失関数の最小化問題となるが、これを解析的に解くことは困難であるため、一般的なニューラルネットワークの学習に用いられる数値的解法により反復的にパラメータを更新することで近似的な解を得る。 (7) 式を用いた DNN の学習手法を多クラス識別へ応用するため、各時点での DNN の出力を N 次元ベクトルとするほか、損失関数を以下のように一般化する

$$\mathcal{L} = \max\left(0, \sum_{k\in\{k_1,k_2\}} \sum_{n=1}^{N} (\phi_{kn} - \psi_{kn} Y_{kn})\right) + \lambda. \tag{9}$$

ここで、$\{k_1,k_2\}\subset\mathcal{K}$ は学習データのインデックスの集合 $\mathcal{K}$ からランダムに選択されたインデックスの対である。ϕ_{kn} および ψ_{kn} は N クラス識別のために導入されたラベルであり、以下のように定めた

$$\phi_{kn} = \begin{cases} 1 & \text{if data belongs to } n\text{th class,} \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

$$\psi_{kn} = \begin{cases} 1 & \text{if data belongs to } n\text{th class,} \\ -1 & \text{otherwise.} \end{cases} \tag{11}$$

また、正則化項 λ は

$$\lambda = \lambda_{smooth} + \lambda_{sparse} + \lambda_{entropy} \tag{12}$$

$$\lambda_{smooth} = p_1 \sum_{k=1}^{2} \sum_{n=1}^{N} \sum_{t=1}^{T-1} (y_{kn}^t - y_{kn}^{t+1})^2, \tag{13}$$

$$\lambda_{sparse} = p_2 \sum_{k=1}^{2} \sum_{n=1}^{N} \sum_{t=1}^{T} y_{kn}^t, \tag{14}$$

$$\lambda_{entropy} = -p_3 \sum_{k=1}^{2} \sum_{n=1}^{N} \phi_{kn} \log \frac{exp(Y_{kn})}{\sum_m exp(Y_{km})} \tag{15}$$

であり、1 項目、2 項目は (8) 式と同様にそれぞれ平滑化項、およびスパース化項である。3 項目はクロスエントロピー損失項であり、DNN の学習を安定させるために導入した。

3.2 DNN を用いた軸受故障検知

提案する DNN の異常検知への適用可能性と識別能力を評価するため、軸受故障診断のための加速度センサデータ [6] を用いた評価を行った。本データセットで用いられた軸受には故障を模擬するため、内輪(IR)、転動体(B)および外輪の 6 時方向、3 時方向および 12 時方向(OR@6,3,12) に放電加工を施すことで亀裂が加えられている。振動データは、モータに取り付けられた加速度計から取得されている。実験では図 4 に示すような DNN を設け、(12) 式の損失関数のハイパパラメータを $p_1=10^{-1}$、$p_2=10^{-5}$、$p_3=10^{-1}$、学習係数 10^{-3} とし、Adam optimizer を用い、(9) 式を最小化するように 10,000 回のパラメータ更新を行った。

図 5 に、それぞれの振動データおよび推定値を示す。異常に特徴的な箇所に対して高い値が推定され、その他の箇所に対しては低い値が推定された。この結果は提案

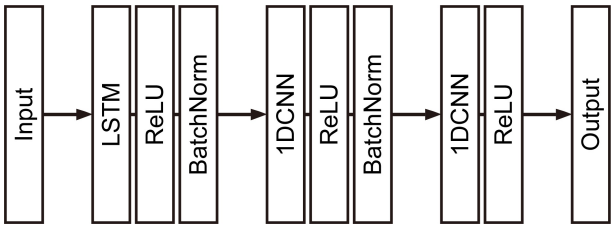

図４　多変量時系列解析のための DNN の構造

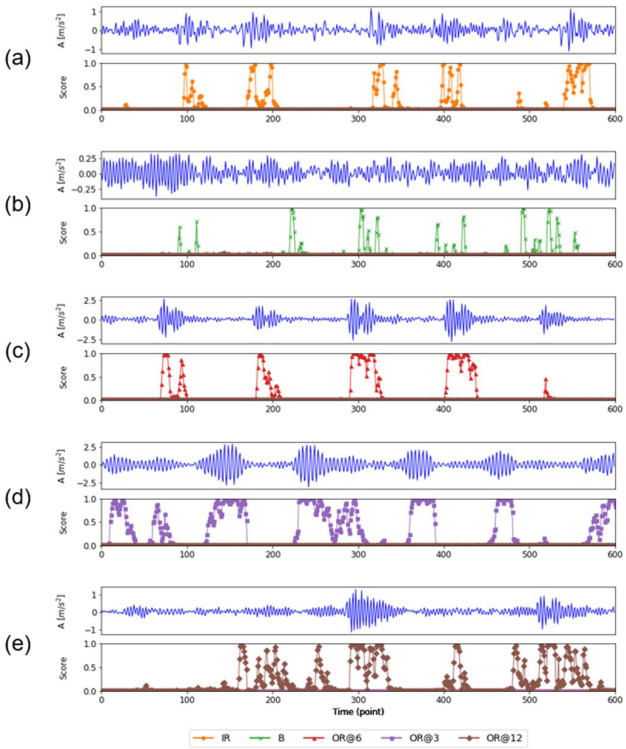

図５　加速度センサデータおよび推定された異常値
（a：IR、b：B、c：OR@6、d：OR@3、e：OR@12）

する DNN および学習方法によって、データに潜在する特徴を自動的に抽出および異常の識別ができることを示している。

3.3 DNN を用いた軸受故障検知

上記の DNN を用いて人物動作データからの異常検知が可能であるか確認した。人物動作データには 2.2 と同様に CMU Mocap データ [4] から「Walk」および「Weird Walks」のカテゴリに含まれる歩行動作データをそれぞれ自然および不自然な動作として選出し、学習を行った。評価では同データセットから、自然な歩行動作（Subject 91, trial 2）および不自然な歩行動作（Subject 91, trial 18）に対し、それぞれの動作に対する異常度を推定した。図 6 に提案手法により推定されたそれぞれの人物動作に対する推定値を示す。不自然な動作を含むデータに対する推定値は、多くの箇所で自然な動作に対する推定値よりも高くなった。さらに、高い値を示す箇所がある一方で、

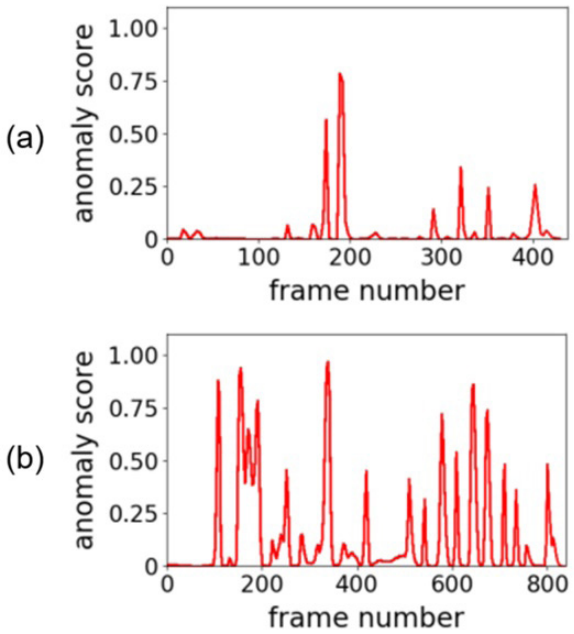

図６　不自然な動作を（a）含まないおよび（b）含む
人物動作データから推定された異常値

低い値を示す箇所も確認された。これは、本手法により学習された DNN が、通常動作からの逸脱を推定した場合にのみ高い値を出力することを示している。

４．おわりに

本稿では、監視カメラ映像からの異常検知を目的とした人物動作特徴量の抽出および多変量時系列データの解析が可能な DNN の学習手法とその評価結果について紹介した。提案する DNN およびその学習手法により、映像中の人物動作の特徴抽出が可能であることを確認した。さらにアノテーションの付与が困難な多変量時系列データから、データの持つ特性を決定付ける上で重要な特徴を自動的に抽出可能な DNN とその学習手法を併せて開発することで、得られた人物動作に関する特徴量から通常時からの逸脱を検知することを可能とした。本技術は原子力関連施設における映像監視等、様々な分野への応用が可能と考えられる。

参考文献

[1] 厚生労働省ホームページ平成 30 年労働災害統計：http://anzeninfo.mhlw.go.jp/user/anzen/tok/anst00.htm

[2]Shi Chen, Kazuyuki Demachi: “A Vision-Based Approach for Ensuring Proper Use of Personal Protective Equipment (PPE) in Decommissioning of Fukushima Daiichi Nuclear Power Station”, Applied Sciences, 10, 5129 (2020)

[3]M. Andriluka, L. Pishchulin, P. Gehler, B. Schiele: “2D

human pose estimation: new benchmark and state of the art analysis”. In: Proceedings of the IEEE Conference on computer Vision and Pattern Recognition, pp. 3686–3693 (2014)

[4]CMU Graphics Lab Motion Capture Database: http://mocap.cs. cmu.edu/

[5]W. Sultani, C. Chen, and M. Shah. “Real-World Anomaly Detection in Surveillance Videos”. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6479–6488 (2018)

[6]Case Western Reserve University Bearing Data Center: http://csegroup.case.edu/bearingdatacente-r/home

（2020 年 11 月 6 日）

著 者 紹 介

著者：三木　大輔
所属：地方独立行政法人東京都立産業技術研究センター，東京大学
専門分野：機械学習，コンピュータビジョン