頭部・眼球運動を用いた深層教師なし学習による個人差に不変な行動可視化手法の提案

山下　純平† 瀧本　祥章†† 小矢　英毅† 大石　晴夫† 熊田　孝恒†††

† 日本電信電話株式会社　NTT アクセスサービスシステム研究所　〒 239–0847 神奈川県横須賀市光の丘 1-1
†† 日本電信電話株式会社　NTT サービスエボリューション研究所　〒 239–0847 神奈川県横須賀市光の丘 1-1
††† 京都大学　〒 606–8501 京都府京都市左京区吉田本町
E-mail: †, ††{junpei.yamashita.vd,yoshiaki.takimoto.ar,hidetaka.koya.gk,haruo.oishi.nw}@hco.ntt.co.jp
† † †t.kumada@i.kyoto-u.ac.jp

あらまし　頭部・眼球運動から人間行動を可視化する技術の実現に向けて取り組んでいる．行動を分析する目的で可視化する際には，事前に想定されていない行動種別を可視化できることが重要である．しかし，従来の人間行動認識では，事前に定義したラベルに依存した教師あり学習が用いられるため，ラベルに存在しない行動を認識することができない課題があった．本研究では，深層教師なし学習（深層生成モデル）を用いて，ノイジーな頭部・眼球運動の多次元データを人手による特徴抽出なしに低次元な潜在表現に変換する手法を提案する．評価実験の結果，提案手法によって，分析者は人の目に把握できる低次元空間上で探索的な分析を行うことが可能になり，想定されていない行動種別を含めた行動を可視化できる可能性が示唆された．この際，行動種別ごとにセンサデータが分かれるような次元削減に寄与する新たな正則化を導入することで，そうした次元削減効果だけでなく，さらに学習の安定化や，ジェネレータによって生成されるデータの質の向上といった効果が得られることを確認した．

キーワード　頭部運動，眼球運動，教師なし学習，可視化，潜在空間，深層生成モデル，GAN，敵対的学習，個人差

1 序　　論

様々なセンサを用いて，人間行動を可視化する方法が提案されている．特に，対象場面の限定なしに計測が可能であることから，装着型センサを利用した方法が広く普及している [24]．この方法では，慣性センサ（加速度計，角速度計）を用いて計測された身体運動から，行動（着座，歩行，階段の上り下りなど）を推定することが一般的である [24]．しかしながら，慣性センサによる方法では，人間が多くの時間を費やしている，目立った身体運動を伴わない認知的行動（会話，読書，PC 作業など）を取得できない課題があった．しかし，このような身体運動を伴わない行動も，殆どの場合，視覚による情報収集を伴っている [1] ことを指摘できる．すなわち，注視位置を左右する頭部・眼球運動の特徴は，身体運動を伴わない行動であっても，その内容を反映して変化すると考えられる．そこで我々は，頭部・眼球運動を用いた幅広い行動可視化の実現を目指している．

近年の典型的な人間行動認識（Human Activity Recognition: HAR）では，事前に訓練データを取得し，センサデータを入力，行動ラベルを正解データとして，深層学習の枠組みで教師あり学習を行う [24]．その後，適用先場面で取得されたセンサデータを入力することで，行動ラベルの推定を行う．このような深層教師あり学習の枠組みでは，大規模なラベル付きデータセットを訓練に用いることができれば，様々な場面で適用可能な認識モデルを得られる可能性がある．しかしながら，HAR においては，大規模なラベル付き公開データセットが存在しない課題がある．これは，日常生活中の行動ラベルを取得することが非常に困難であるためと考えられる．このようなデータセットを得るためには，センサを装着して日常生活を多数の個人に送ってもらい，かつ，その様子を常に撮影するなどして，時刻ごとの行動ラベルを付与する必要がある．しかし，コストの観点だけでなく，プライバシの観点からも，こうした大規模データを収集，公開することは困難である．そこで，多くの HAR の研究においては，実験室内で，実験参加者に日常生活を模擬した行動を行ってもらい，その際に取得したラベルを訓練データに用いて行動を認識する代替的なアプローチがとられる [24]．ところが，このように規模がそれほど大きくないデータセットを用いた教師あり学習では，訓練データとラベルの分布と，推定データとラベルの分布を一致させるよう注意を払わなければ，モデルによる推定は有効に機能しない [4]．最もシンプルな例では，実験室環境においてラベルとして定義されていない行動種別が日常場面に存在する場合があげられる．この場合，モデルはその定義されていない行動種別について，全て誤った認識結果を出力することになる．しかし，実験室内環境で模擬できる行動種別には限りがあるため，HAR は原理的にこの問題を回避できない．

一方，オフィス業務における従業員の行動を振り返る場合など，行動種別を分析的に取得したい場合には，このような「事前に想定されていない行動種別」を可視化できることは非常に重要である．そこで，本研究では，ラベルが付与されていない適用場面のデータでモデルを訓練することができる，教師なし学習を用いた行動可視化手法を提案する．教師なし学習では，

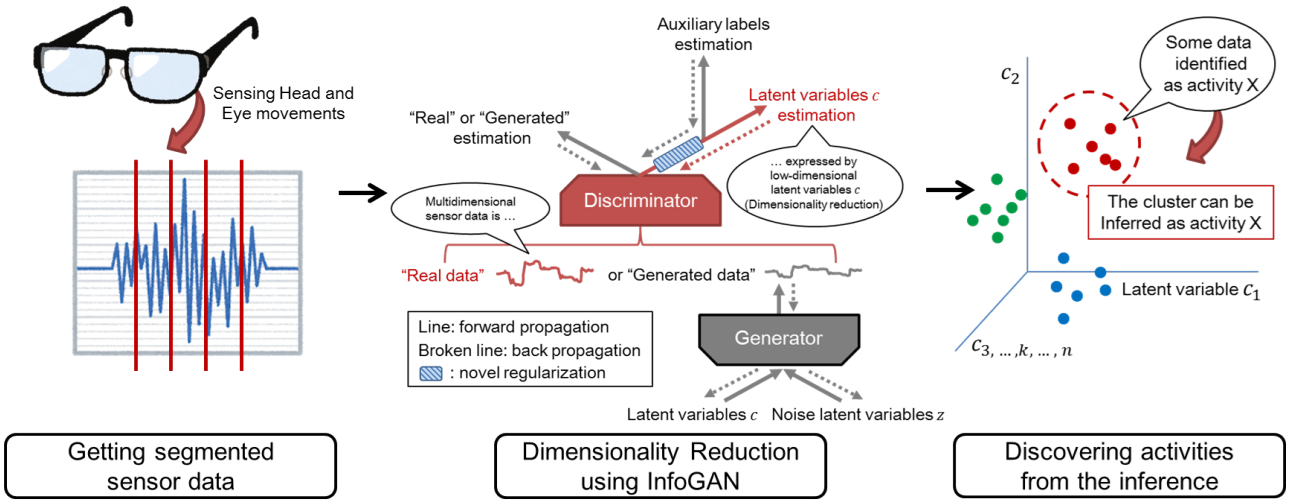

図 1　提案手法による行動可視化の流れ

多次元データを人の目に把握可能な程度まで次元削減し，得られた低次元の表現を事後的に分析することで，データに関する知見が得られる [22].

本研究では，このような次元削減を行うために，相互情報量最大化敵対的生成ネットワーク（Information Maximizing Generative Adversarial Networks: InfoGAN）を用いる方法を提案する（図 1）．InfoGAN によって，低次元な潜在空間上に分布する潜在変数から，多次元の頭部・眼球運動のセンサデータが生成される過程をモデル化することで，センサデータを，そのデータを生成した少数の潜在変数の組によって表現でき，大幅な次元削減が可能になる．その後，分析者は，いくつかのセンサデータの行動種別を事後的に調査することで，その周囲のデータも同じ行動種別を意味していると推量することができる．こうした流れで，行動の可視化が実現される．

上記のことから，提案手法による行動可視化では，次元削減が行われる際，行動種別の観点から，類似したデータ同士は近くに分布し，異なったデータ同士は分かれて分布することが望ましい．しかし，既存の InfoGAN では，全ての顕著な特徴のばらつきを潜在空間上の次元に対応させるため，センサデータは行動種別のばらつきだけでなく，個人差などのばらつきにも対応して潜在空間上に分かれて分布してしまう．そこで，我々は，望んだ説明のばらつきのみを潜在空間の次元によって説明する新たな正則化アーキテクチャを InfoGAN に導入する．この正則化によって，潜在空間上において，センサデータは行動種別ごとに分かれて分布するようになり，低次元表現を用いて行動を可視化することが容易になると期待される．なお，本研究にあたっては，予備的な実験検討を [25] にて行っている．

2　関 連 研 究

2.1　頭部・眼球運動計測

頭部運動は，慣性センサを装着することで，容易に日常的な計測が可能となる．一方，眼球運動の計測はそれほど容易ではない．注視位置を取得したい場合に一般的に用いられるカメラを用いた方法では，高精度に眼球運動が取得できるが，大容量データを高速で処理する必要性から，日常的に装着して計測を行うことは容易ではない [10]．カメラを用いずに眼球運動を計測する方法として，眼電位（Electrooculography: EOG）を用いた方法がある．EOG では．眼球が前方に＋，後方に－の電位を持っていることを利用して，眼球の周囲に電極を貼ることにより，眼球の相対的運動を推定できる [10]．既存研究において，Ishimaru ら [10] は，慣性センサおよび電極を装着したメガネ型ウェアラブルデバイス JINS MEME によって，広い場面で頭部・眼球運動を計測可能であり，その信号は一定程度ノイジーではあるものの，実験室内データにおける HAR には十分であることを示している．

2.2　頭部・眼球運動による行動可視化

頭部，眼球運動に着目した行動可視化手法のうち，眼球運動を用いた手法については，多くの既存研究がある．これらでは，多くの場合，簡便に計測できる EOG が用いられる．Bulling ら [1] は，眼電位から，サッカード（急速眼球運動）を，人手によって（ルールベースのアルゴリズムなどで）事前抽出し，6 種類の行動分類を行っている．JINS MEME によって計測された眼球運動を用いた研究には，[5,10] がある．Ishimaru ら [10] は，EOG 信号を統計量に要約し，4 種類の行動分類を行っている．Díaz ら [5] は，より日常に近いスマートホーム環境において，サッカードなどを人手によって事前抽出し，12 種類の行動分類を行っている．これらの研究は実験室内データを対象にした教師あり学習による HAR であるため，日常生活におけるセンサデータを対象として，ラベルが定義されていない行動種別を認識することはできない．

数少ない，日常生活中の行動データを用いた眼球運動による行動可視化手法の提案に，Steil ら [22] の報告がある．Steil らは，被験者にカメラベースの眼球運動計測装置を装着して生活を送ってもらい，人手によって事前抽出したサッカード，注視，瞬目などを特徴量として教師なし学習を行うことで，いくつかの顕著な行動を発見できることを報告している．Steil らの手法

は教師なし学習を用いているため，ラベルが定義されていない行動種別を認識できない課題は生じないが，高精度な信号からサッカードなどの特徴量を事前抽出することが必要であり，ノイジーな EOG 信号に適用することは困難である．

頭部運動を用いた手法については，眼球運動ほど多くの例は見当たらない．教師あり学習による HAR の例として，Tan [23] らは，カメラを用いて対象者を撮影し，頭部運動の特徴から 10 種類の行動分類を行っている．Madabhushi [17] らは，同様にカメラによって計測された頭部運動から 12 種類の行動分類を行っている．これらの研究も，ラベルが定義されていない行動種別を認識することはできない．また，装着型のセンサを用いていないため，対象者が限定された撮影範囲内に存在している必要がある．

頭部運動，眼球運動の両者を用いた手法も同様に多くはない．Ishimaru ら [9] は，加速度センサを用いて頭部運動を取得し，眼球運動から事前抽出した瞬目頻度と組み合わせて用いることで，5 種類の行動分類を行っている．彼らの研究では，瞬目頻度に頭部運動を組み合わせることで，分類精度が大きく向上することが示されている．この研究も，ラベルが定義されていない行動種別を認識することはできない．また，Ishimaru らは眼球運動のうち，瞬目のみを事前抽出して入力に用いているが，本研究ではそうした事前抽出による情報の削減なしに眼球運動を用いる点が異なる．

2.3 深層教師なし学習（深層生成モデル）

近年，深層学習を用いた教師なし学習によって，多次元データを人の目に把握できる程度の低次元空間に写像する研究が進展している．深層生成モデルを用いた手法では，潜在変数の組から多次元データが生成される過程をモデル化することで，次元削減を実現できる．具体的なアルゴリズムとしては，潜在変数から多次元の観測データを生成するデコータと，観測データから潜在変数を推定するエンコーダからなる変分オートエンコーダ（Variational Autoencoders: VAE）[15] や，潜在変数から多次元の観測データを生成するジェネレータと，生成された観測データと実データを識別するディスクリミネータからなる敵対的生成ネットワーク（Generative Adversatial Networks: GAN）[7] などがある．VAE では，多次元の観測データをエンコーダに通すことで，データ生成の元となる潜在変数が推定される．GAN では，観測データから潜在変数を推定する経路を追加したディスクリミネータを用いることで，潜在変数が推定される [3]．これらの方法によって，多次元データを人の目に把握できる程度の低次元空間に写像することが可能となる．しかしながら，我々の知る限り，頭部・眼球運動に関連する信号を深層生成モデルによって特徴抽出なしに低次元空間上に表現した試みは存在しない．

潜在変数によって生成データの特徴を説明する際，特徴のばらつきを解釈しやすい形で潜在空間の次元に対応させることについては，主に Disentanglement と呼ばれる技術分野で研究が盛んである [13]．しかし，Disentanglement では，すべての特徴のばらつきを潜在空間の互いに独立した次元に対応させることを目指す [13]．一方，本研究では，望んだ特徴のばらつきのみを潜在空間の（独立した）次元に対応させることを目指す．

3 提 案 手 法

提案手法では，InfoGAN を用いて，JINS MEME[1]によって取得されたノイジーな頭部・眼球運動のセンサデータの生成過程を，人手による特徴抽出なしでモデル化する．このモデルを利用することで，センサデータを，そのデータを生成した少数の潜在変数の組によって表現でき，大幅な次元削減が可能になる．この際，望んだ説明のばらつきのみを潜在空間の次元によって説明する新たな正則化アーキテクチャを導入することで，潜在空間上において，センサデータは行動種別ごとに分かれて分布するようになる．これにより，分析者は，いくつかのセンサデータの行動種別を事後的に調査することで，その周囲のデータも同じ行動種別を意味していると推量することが可能になる．以下では，まず，準備として提案手法に用いている GAN および InfoGAN と，新たな正則化に用いている Gradient Reversal Layer について説明する．次に，提案手法と正則化アーキテクチャの詳細について説明する．

3.1 準　　備

3.1.1 GAN

GAN [7] では，ディスクリミネータは，自身に入力されるデータがジェネレータが生成したデータであるか実データであるかを識別するように学習していく．一方で，ジェネレータは，自身が生成したデータがディスクリミネータによって識別されないように学習していく．このようなジェネレータとディスクリミネータの敵対的学習によって，ジェネレータには潜在変数からデータが生成される過程がモデル化される．

3.1.2 InfoGAN

InfoGAN [3] は，GAN のディスクリミネータに，潜在変数の推定を行う経路を追加したものである．InfoGAN では，ジェネレータは生成過程の元となる潜在変数 c，ノイズ潜在変数 z を入力として，実データに類似したデータの生成を行う．ディスクリミネータでは，実データと生成データの識別に加えて，潜在変数 c の推定を行う．潜在変数 c の推定誤差をジェネレータに誤差として加えることで，ジェネレータは，ディスクリミネータが潜在変数 c を推定しやすいように出力を生成するようになる（生成データと潜在変数 c の相互情報量が最大化される）．これは，潜在変数 c の違いによって，生成されるデータの特徴ができるだけ顕著に異なるように生成過程がモデル化されることを意味している．上記の学習は，以下の式で説明される（x：観測（実）データ，G：ジェネレータ，D：ディスクリミネータ，I：相互情報量，λ：潜在変数 c と生成データの相互情報量最大化に関する重み），

$$\min_D \max_G \log D(x) + \log(1 - D(G(z, c))) + \lambda I(c; G(z, c)).$$

式の第 1 項は，ディスクリミネータが実データを入力された

1：https://jins-meme.com/ja/

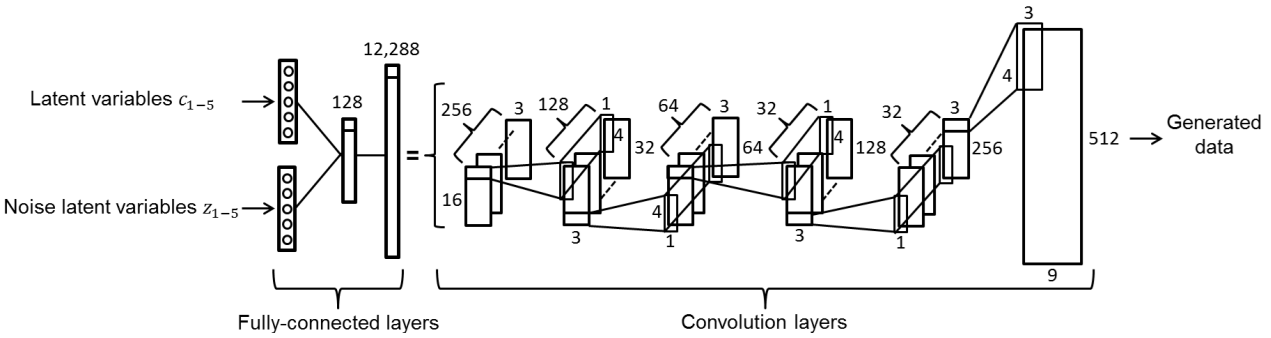

図 2　提案手法のアーキテクチャ（ジェネレータ）

場合に最小の値を出力するように学習することを，第 2 項は，ディスクリミネータが生成データが入力された場合に最大の値を出力するように学習することを，第 3 項は，ジェネレータが潜在変数 c と生成したデータの相互情報量が最大になるように学習することを表している．

3.1.3　Gradient Reversal Layer

Gradient Reversal Layer（GRL）とは，入力と同数の出力を返す学習対象となるパラメータを持たない層であり，学習時に符号を逆にした勾配を逆伝播することを特徴とする．GRL は，一般にはドメイン適応を実現するために用いられる．ドメイン適応とは，十分な教師ラベルを持つドメイン（ソースドメイン）から得られた知識を，十分な情報がないドメイン（ターゲットドメイン）に適用するための技術である（ここで，ドメインとは，データの集まりを指す）．Ganin ら [6] は，ラベルを推定するモデルの教師あり学習を行う際に，入力データが属するドメインを識別する 2 層以上の経路を中間層に追加し，追加した経路の 1 層目に GRL を用いた．この経路について誤差逆伝播法によって学習を行うと，GRL 以降ではドメイン識別の精度が向上するよう重みが学習されていくが，GRL 以前では逆転された（マイナスの値で乗算された）勾配が伝播していくため，ドメイン識別の精度が低下するよう重みが更新されていく．これによって，この経路が接続されている以前の層においては，ドメインの識別に寄与し得る情報を喪失させるような正則化効果が得られる．ソースドメインにおいて教師あり学習を行いながら，このような正則化を与えることで，ソースドメインとターゲットドメインに共通してラベル推定が行えるモデルを得ることができ，ドメイン適応が可能になると示されている．

3.2　提案手法のアーキテクチャ

提案手法のアーキテクチャを図 2（ジェネレータ），図 3（ディスクリミネータ）に示した．ジェネレータは，2 つの全結合層と 5 つの逆畳み込み層で構成されている．潜在変数の次元数と，ノイズ潜在変数の次元数を加えた長さの配列からなる入力データは，ユニット数 128 の全結合層に入力され，Relu function で活性化が行われる．次に，ユニット数 12,288 の全結合層に入力され，Relu function で活性化が行われる．その後，値は逆畳み込み層を経て多次元（9×512）の生成センサデータに変換される．1 層目から 4 層目までの逆畳み込み層は，センサ種別方向に幅 1，時系列方向に幅 4 のカーネルであり，ストライドはセンサ種別方向に 1，時系列方向に 2 である．5 層目ではセンサ種別方向に幅 3，時系列方向に幅 4 のカーネルであり，ストライドはセンサ種別方向に 3，時系列方向に 2 である．頭部加速度，頭部角速度，眼電位の各センサは 3 次元ずつの信号であるため，カーネルはセンサ種類をまたがないように適用される．各層におけるカーネル数は，1 層目で 128 個，2 層目で 64 個，3 層目で 32 個，4 層目で 32 個，5 層目で 1 個である．全ての層で，パディングはセンサ種別の方向に 0，時系列方向に 1 である．中間層の逆畳み込み層では，入力信号とカーネルの逆畳み込みを計算したのち，Leaky Relu function で活性化が行われる．なお，本アーキテクチャにおける Leaky Relu function では，入力値域が 0 以下の場合 0.2 が乗算される．出力層の逆畳み込み層では，シグモイド関数で活性化が行われる．出力層を除き，Batch Normalization [8,19]，および Dropout [11,21]（rate $= 0.5$）を用いる．

ディスクリミネータは，全ての経路に共通した，5 つの畳み込み層と 1 つの全結合層の後に，経路ごとに異なる 1–5 層の全結合層で構成されている．畳み込み層の後，生成データと実データの識別を行う経路では，1 つの全結合層が構成されている．潜在変数の推定を行う経路では，3 つの全結合層が構成されている（InfoGAN [3] では，この経路の層数は 2 であるが，新たな正則化による特徴喪失効果を十分に得るため 3 としている）．新たな正則化を行うために個人を推定する経路では，潜在変数の推定を行う経路と 2 層の全結合層を共有した後，1 つの GRL の後，2 つの全結合層が構成されている．この経路による正則化については後述する．まず，センサ種別方向に 9 次元，時系列方向に 512 次元の入力データは，畳み込み層を通過する．本モデルの畳み込み層は，1 層目ではセンサ種別方向に幅 3，時系列方向に幅 4 のカーネルであり，ストライドはセンサ種別方向に 3，時系列方向に 2 である．頭部加速度，頭部角速度，眼電位の各センサは 3 次元ずつの信号であるため，カーネルはセンサ種類をまたがないように適用される．2 層目から 5 層目までの畳み込み層は，センサ種別方向に幅 1，時系列方向に幅 4 のカーネルであり，ストライドはセンサ種別方向に 1，時系列方向に 2 である．各層におけるカーネル数は，1 層目で 32 個，2 層目で 64 個，3 層目で 128 個，4 層目で 256 個，5 層目で 512 個である．全ての層で，パディングはセンサ種別の方向に 0，時系列方向に 1 である．これらの畳み込み層では，

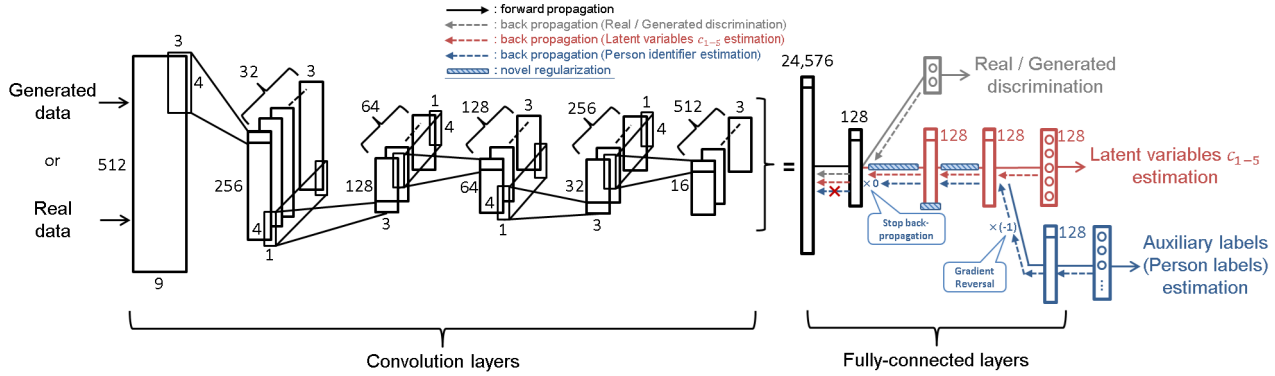

図 3　提案手法のアーキテクチャ（ディスクリミネータ）

カーネルの畳み込みを計算したのち，Leaky Relu function で活性化が行われる．そののち，値はユニット数 128 の全結合層に入力され，Relu function で活性化が行われる．生成データと実データの識別を行う経路では，そこから 2 ユニットからなる出力層に入力される．潜在変数の推定を行う経路では，そこから値は 128 ユニットからなるもう 2 つの中間層に入力され，それぞれ Leaky Relu function で活性化が行われた後，潜在変数 c の次元数分のユニットからなる出力層に入力される．新たな正則化を行う経路では，潜在変数の推定を行う経路と 2 層の全結合層を共有した後，値は GRL に入力される．GRL は，順伝播時には入力をそのまま出力する．その後，128 ユニットからなるもう 1 つの中間層に入力され，Leaky Relu function で活性化が行われた後，データに含まれる個人数分のユニットからなる出力層に入力される．生成データと実データの識別を行う経路の出力層を除き，全ての層で Spectral Normalization [18] を用いる．

3.3　新たな正則化アーキテクチャ

本提案手法では，InfoGAN のディスクリミネータに個人を識別する経路を追加し，GRL によって個人差を打ち消す学習を行う．具体的には，潜在変数 c の推定を行う出力層が分岐している層（図 3 の赤い経路）に，個人を識別する 3 層の経路（青い経路）をさらに追加し，その 1 層目に GRL を用いる．この経路で個人の識別精度が向上していくよう学習を進めると，GRL より前に位置する潜在変数 c の推定経路（赤い経路のうち，青い斜線の塗りつぶしで示された部分）においては，入力された特徴量から個人の識別に寄与しうる情報を打ち消して出力するような学習が行われいく．ただし，個人識別に関する誤差勾配の逆伝播は，実データ/生成データの識別が行われている出力層が合流するユニット数 128 の層に至る前で停止させる．これにより，実データ/生成データの推定は，個人差に関する特徴を考慮して行われるが，潜在変数 c の推定は，個人差に関する特徴のみを選択的に考慮せずに行われるようになる．これによって，もし，潜在変数 c が，データのばらつきのうち個人差に起因するものを表現していると，その差はディスクリミネータ内で推定不能となり，潜在変数 c に関する推定誤差を増大させることとなる．ジェネレータにおいては，ディスクリミネータの潜在変数 c に関する推定誤差を最小化するように学習が進められるため，潜在変数 c には個人差が反映されないように生成過程がモデル化される．一方，ディスクリミネータは個人差も含めて，生成データが実データと類似していることを要求するため，結果的にジェネレータは個人差をノイズ潜在変数 z によって説明するように学習を行っていく．この過程の導入によって，提案手法の式は以下のようになる（λ_1：潜在変数 c と生成データの相互情報量最大化に関する重み，λ_2：個人差と潜在変数 c の相互情報量最小化に関する重み，個人ラベル：p），

$$\min_D \max_G \log D(x) + \log(1 - D(G(z,c))) \\ +\lambda_1 I(c; G(z,c)) - \lambda_2 I(p; G(z,c) \mid z).$$

式の第 1–3 項は，既存の InfoGAN と同様である．第 4 項は，ノイズ潜在変数 z を固定した場合，ジェネレータが，個人ラベルと生成したデータとの間の相互情報量が最小化するように学習することを表している．

4　評 価 実 験

4.1　方　　法

本実験では，提案手法を評価するため，オフィス内を再現した 4 種類の行動中のセンサデータを取得し，次元削減による可視化を行った．具体的には，実験参加者 14 名に契約書類（見積書，発注書など）からデスクトップ PC へのデータ入力，スマートフォンの利用，紙文書の筆記具を用いた校閲，実験者との会話（実験についての感想など）を 10 分間行ってもらい，その際のセンサデータを計測した．メガネ型ウェアラブルデバイス JINS MEME ESR[2]を用いて，頭部加速度 3 軸，頭部角速度 3 軸，水平・垂直方向の眼電位差，および左側電極の眼電位を 50Hz で計測した．

4.2　データ前処理

提案手法では，ジェネレータが [0, 1] の範囲で出力を行うシグモイド関数による出力層を備えているため，実データも同様に [0, 1] の範囲に収める必要がある．そこで，ほぼすべてのデータが範囲内に収まるよう，センサデータは，それぞれの平

2：https://jins-meme.com/ja/researchers/specifications/

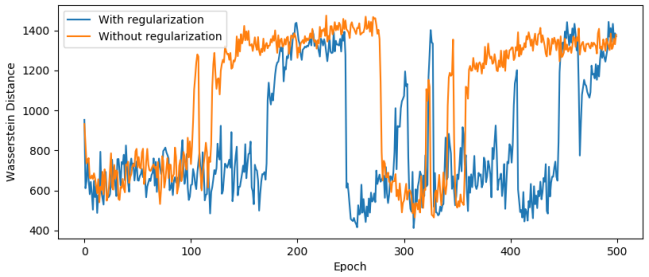

図 4　両モデルについての学習エポックごとの WD

表 1　両モデルの上位 10 位までの Wasserstein Distance

Regularization	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
With	**412**	**416**	**433**	**443**	**443**	**445**	**447**	**449**	**452**	**457**
Without	465	465	474	478	486	490	490	505	513	514

表 2　両モデルの WD が上位 10 位までのエポックにおける F1 score

Regularization	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
With	.647	**.684**	**.683**	**.689**	**.686**	.644	**.690**	.640	**.691**	**.692**
Without	**.663**	.651	.648	.652	.663	**.660**	.673	**.660**	.668	.648

均値を引いたのち，頭部運動に関するデータは 50,000，眼電位に関するデータは 5,000 で割り，そして 0.5 を加算された．範囲外の値を持つ数点のデータは除外された．

その後，センサデータは 10.24 秒区間の移動窓で分割した（窓の長さは予備実験における試行錯誤によって決定した）．これにより，センサ種別方向に 9 次元，時系列方向に 512 次元の入力データが得られた．移動窓は時系列上を 64 ポイント単位で動く基準から，個々の窓において $[+0, +32]$ ポイントの範囲でランダムにずれたポイントでデータを取得した．

4.3　学　　習

ジェネレータの入力である潜在変数 c，ノイズ潜在変数 z の次元数はそれぞれ 5 に設定された．潜在変数 c は平均値 0，標準偏差 1 の正規分布からサンプリングされた．ノイズ潜在変数 z は $[-1, 1]$ の範囲を取る一様分布からサンプリングされた．

ジェネレータの誤差関数には，ディスクリミネータの生成データと実データの識別について，交差エントロピー関数の負の出力を用いた．ジェネレータの学習について，最適化には Adam [14] （$\alpha = 0.001, \beta_1 = 0.5$）を用いた．

ディスクリミネータの誤差関数には，ディスクリミネータの生成データと実データの識別について，交差エントロピー関数を用いた．個人の識別については，交差エントロピー関数の出力に重み $\lambda_2 = 0.1$ を乗算して用いた．最適化には Adam [14]（$\alpha = 0.002, \beta_1 = 0.5$）を用いた．

潜在変数 c の推定については，潜在変数 c の推定値を平均としたガウス分布（分散は微小値）の負の対数尤度関数に，実際の潜在変数 c を入力した結果を重み $\lambda_1 = 0.25$ を乗算し，ジェネレータ，ディスクリミネータの誤差に加えた．

個人差を打ち消す正則化の効果を検証するため，64 データごとのミニバッチ単位で，正則化あり，なしのモデルをそれぞれ 500 エポック学習した．なお，正則化なしモデルでは，潜在変数 c の推定経路は InfoGAN 元論文 [3] と同様に 2 層とした．

表 3　典型データの変化量の総和の平均値

Movements	Proofread	Smartphone	Data entry (PC)	Talk
Head (acceleration)	17.12 (±0.02)	13.68 (±0.01)	11.27 (±0.01)	11.18 (±0.01)
Head (rotation)	8.88 (±0.02)	9.11 (±0.02)	9.44 (±0.03)	9.78 (±0.03)
Eye (horizontal)	11.62 (±0.10)	8.09 (±0.03)	8.08 (±0.01)	7.95 (±0.01)
Eye (vertical)	17.27 (±0.10)	12.46 (±0.04)	12.30 (±0.03)	12.31 (±0.01)

4.4　結　　果

正則化あり，正則化なしの両モデルについて，学習エポックごとにジェネレータが生成するセンサデータの質を評価した．画像データを対象とした GAN の評価指標に，Sliced-Wasserstein Distance（SWD）[12] がある．SWD では，ジェネレータが生成したデータと実データの分布の差を距離として指標化することで，ジェネレータが実データに似たデータを生成できていることを検証する．本研究では，2 次元の画像データの分布間距離を測る SWD の算出方法に則り，1 次元のセンサデータの分布間距離を Wasserstein Distance（WD）として算出し，各エポックにおける生成データの質を評価した結果を図 4 に示した（算出時のパラメータは，切り出されるミニバッチが 1 次元方向のみに伸びていることを除いて [12] に則った）．正則化なしのモデルは，学習が崩壊し著しく大きい WD が算出されているエポック数が多いが，正則化ありのモデルでは，そのようなエポックが少なく，学習が安定していることがわかる．両モデルのエポック中で上位 10 番目までの WD を表 1 に示した．正則化ありのモデルは，正則化なしのモデルに比べて，WD が良好なエポックにおけるデータの質も上回っていることがわかった（t 検定により有意，$t(18) = -6.53, p < .001$）．すなわち，正則化によって，学習が安定化し，かつジェネレータが生成するデータの質も向上することがわかった．

次に，ディスクリミネータを用いて，センサデータを，データを生成する元である 5 次元の潜在変数 c の組で表現した結果を分析した．まず，比較的初期である 50 エポック目におけるディスクリミネータによって得られた潜在空間上のセンサデータ分布を図 5，6 の左側に示した（5 次元の潜在空間を，t-SNE [16] を用いて 2 次元上に描写している）．正則化なしのモデルにおいては，全ての行動種別について，一部のセンサデータが，個人差に影響され，他の行動種別と混ざって分布していた．一方，正則化ありのモデルにおいては，そうした個人差が打ち消され，センサデータは行動種別ごとにまとまって分布していた．

次に，学習エポック全体を対象として，実際の行動可視化場面を想定した分析を行った．実際の可視化場面においては，行動種別ラベルごとに分かれた潜在空間が得られているかを参照してから，良いモデルを選択することはできない．代わりに，WD の値が良いエポックのモデルを用いて分析を行うことを想定した．正則化あり，正則化なしの両モデルについて，それぞれの上位 10 位に入る WD が計測されたエポックにおけるディスクリミネータを抽出し，それらによって得られた潜在空間上のセンサデータ分布を図 5，6 の右側 に示した（同様に t-SNE [16] を用いた）．正則化なしのモデルにおいては，一部の文書校閲，スマートフォン利用中のセンサデータが，依然として個人差に影響され，ばらついて分布していることが見て取れる．一

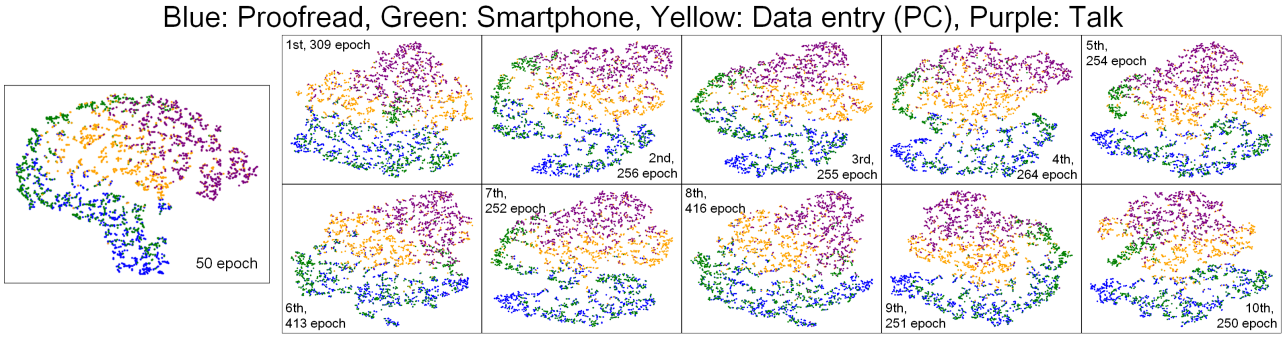

図 5　正則化ありモデルの 50 エポック目，および WD 上位 10 位までのエポックにおける，潜在空間上でのセンサデータの分布

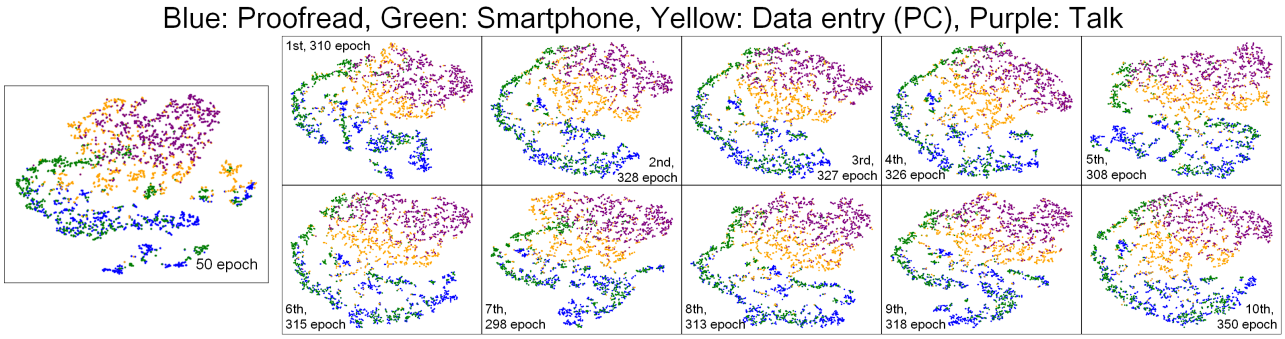

図 6　正則化なしモデルの 50 エポック目，および WD 上位 10 位までのエポックにおける，潜在空間上でのセンサデータの分布

方，正則化ありのモデルにおいては，そうした個人差が打ち消され，センサデータは行動種別ごとにまとまって分布していた．

これらの空間上において，センサデータが行動種別ごとに分かれている境界の明瞭さを定量的に評価した．具体的には，センサデータを潜在空間上に表現した後，Support Vector Machine（SVM） [2] によって教師あり学習を行い，その精度を One-person-leave-out-cross validation によって算出される分類精度（F1 score）で評価した（パラメータ C, γ には，0.1, 1, 10 のいずれかを用い，最も精度の良かったものを採用した）．この方法によって算出される分類精度は，センサデータが行動種別ごとに分かれているが，個人差によってはばらついていない場合に向上する．結果を表 2 に示した．正則化ありのモデルは，正則化なしのモデルに比べて，分類精度が向上していることから（t 検定により有意，$t(18) = 2.18, p < .05$），正則化によって行動可視化を容易にする次元削減が行われていることが確認された．

なお，WD が最も良いエポックの F1 score は低下していたことから（表 2），ジェネレータの質が最良のモデルが，ディスクリミネータによる次元削減においても優れているとは限らないこともわかった（F1 score = .647）．本研究での学習結果を参照する限り，WD と F1 score の両方で優れたモデルを選定するには，前後エポックを含めた平均 WD が最も良い安定的なモデルを選ぶことが有効であった（この方法を用いると，正則化ありの場合，WD 上位 3 位，F1 score = .683 のモデルが選定される）．

最後に，われわれは，分析者が行動種別の特徴をセンサデータの特徴から分析する場合を想定した評価を行った．具体的には，先の基準で選定されたエポックのディスクリミネータによって算出される潜在空間上における各行動種別の中心座標を，典型的なデータが存在する位置とみなして，その位置に潜在変数 c を固定し，ノイズ潜在変数は学習時と同様の条件でサンプリングして，ジェネレータによって典型的なデータを 12,800 点生成し，その特徴を調べた．各行動種別の典型データ（10.24 秒間，値は [0, 1] 間に正規化）の変化量の絶対値を総和を算出した結果の平均を表 3 に示した（上から順に 2 行目から，加速度，角速度，水平眼球運動，垂直眼球運動に対応した総和の平均を記載．括弧内は標準偏差を記載）．これら変化量を，頭部や眼球の動きの量として解釈した．なお，頭部加速度には，加速度センサが重力を加速度として記録するため，頭部が向いている絶対的な方向，すなわち姿勢を反映した成分が多く含まれており，その変化量は姿勢変化量を反映したと考えられる．各行動について変化量が特徴的だったものについて記載する．まず，文書校閲，スマートフォン利用において，頭部方向（姿勢）の変化量が多かった．これらは，注視対象が比較的軽量で取り回しのきく物体（紙文書，スマートフォン）である際に，様々な姿勢がとられたことを示唆している．次に，データ入力，会話において頭部角速度の変化量が多かった．データ入力では，ディスプレイと卓上の書類間を往復する大きな注視位置の移動時に，頭部方向の回転が発生したことが考えられる．会話では，頷きなどのジェスチャが生じていた可能性がある．眼球運動に

ついては，文書校閲において大きな変化量が見られた．一方で，頭部角速度の変化量は小さかったことと合わせて考えると，実験参加者は，紙書類が対象の場合には，頭部運動ではなく眼球運動のみで注視位置を移動させる傾向があったことを示唆している．最も小さい眼球運動の変化量が観測されたのは，会話中であった．これは，会話の相手に視線を向け続けていたことが理由として考えられる．

5 結　　論

提案手法によって，人手による特徴の事前抽出不要で，教師ラベルの付与されていない頭部・眼球運動のセンサデータを，人に把握できるほど低次元の表現に変換することが可能であった．また，新たに導入した正則化を用いることで，個人差によらず，行動種別ごとに分かれてセンサデータが分布するような次元削減が実現されることが確認された．加えて，新たな正則化は，学習の安定化や，生成データの質向上にも寄与することが示された．さらに，ジェネレータが生成したデータを分析することで，行動種別の特徴について分析することが可能であった．これらの結果は，提案手法によって，日常生活場面においても行動可視化ができる可能性を示唆していた．

残された課題としては，以下のものがある．まず，本研究においては，個人がいくつかのデータについて行動種別を取得できた場合，おおむねその周囲の行動種別を推量できると考察したが，この点についてより定量的な分析が必要である．例えば，具体的にいくつのデータについて行動種別を取得すれば十分な精度で分析が可能であるか，どのような境界を引くと良好な識別が可能について検討が必要である．少数の行動種別をラベルとして用いて，提案手法に半教師あり学習を組み合わせる手法（ss-infoGAN [20]）も有望であるため，検討の必要がある．

次に，GRL による正則化では，個人差が打ち消されるが，これに行動種別の差が含まれうるという課題がある．例えば，いくつかの個人が特異な行動をとっていた場合，その行動の違いは個人差として識別されうるため，そのような行動については，適切な可視化が行えなくなる恐れがある．この問題への対処法としては，Universal Domain Adaptation の考え方が適用できる可能性がある [26]．具体的には，個人の識別性が非常に高いデータは，特異な行動種別に対応する可能性が高いという仮定を置くことで，問題を解決できる可能性がある．これらの課題に対応することで，さらに多様な行動を含む，日常生活でのセンサデータによる行動可視化を実現していく予定である．

文　　献

[1] A. Bulling, J. A. Ward, H. Gellersen, and G., Troster, “Eye movement analysis for activity recognition using electrooculography,” IEEE Trans Pattern Anal Mach Intell, 33(4), pp. 741–753, 2010.

[2] CC. Chang, CJ. Lin, “LIBSVM: A library for support vector machines,” ACM Trans Intell Syst Technol, 2(3), pp. 27:1–27:27, 2011.

[3] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, “Infogan: Interpretable representation learning by information maximizing generative adversarial nets,” NIPS, pp. 2172—2180, 2016.

[4] D. Cook, K. D. Feuz, N. C. Krishnan, “Transfer learning for activity recognition: A survey,” KAIS, 36(3), pp. 537-556, 2013.

[5] D. Díaz, N. Yee, C. Daum, E. Stroulia, and L. Liu, “Activity classification in independent living environment with JINS MEME Eyewear,” PerCom, pp. 1–9, 2015.

[6] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, V. Lempitsky, “Domain-adversarial training of neural networks,” JMLR, 17(1), pp. 2096—2030, 2016.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, and Y. Bengio, “Generative adversarial nets,” NIPS, pp. 2672—2680, 2014.

[8] S. Ioffe, and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” arXiv preprint arXiv:1502.03167, 2015.

[9] S. Ishimaru, K. Kunze, K. Kise, J. Weppner, A. Dengel, P. Lukowicz, and A. Bulling, “In the blink of an eye: combining head motion and eye blink frequency for activity recognition with google glass,” AH, pp. 1—4, 2014.

[10] S. Ishimaru, K. Kunze, Y. Uema, K. Kise, M. Inami and K. Tanaka, “Smarter eyewear: using commercial EOG glasses for activity recognition,” UbiComp, pp. 239–242, 2014.

[11] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” CVPR, pp. 1125–1134, 2017.

[12] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive growing of gans for improved quality, stability, and variation,” arXiv preprint arXiv:1710.10196, 2017.

[13] H. Kim, and A. Mnih, “Disentangling by factorising,” arXiv preprint arXiv:1802.05983, 2018.

[14] D. P. Kingma, and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[15] D. P. Kingma, and M. Welling, “Auto-encoding variational bayes,” arXiv preprint arXiv:1312.6114, 2013.

[16] L. V. D. Maaten, and G. Hinton, “Visualizing data using t-SNE,” JMLR, 9, pp. 2579–2605, 2008.

[17] A. Madabhushi, and J. K. Aggarwal, “Using head movement to recognize activity,” ICPR, pp. 698—701, 2000.

[18] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, “Spectral normalization for generative adversarial networks,” arXiv preprint arXiv:1802.05957, 2018.

[19] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” arXiv preprint arXiv:1511.06434, 2015.

[20] A. Spurr, E. Aksan, and O. Hilliges, “Guiding infogan with semi-supervision,” ECML PKDD, pp. 119—134, 2017.

[21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” JMLR, 15(1), pp. 1929—1958, 2014.

[22] J. Steil, and A. Bulling, “Discovery of everyday human activities from long-term visual behaviour using topic models,” UbiComp, pp. 75—85, 2015.

[23] H. CC. Tan, S. De, and C. Liyanage, “Human activity recognition by head movement using Elman network and Neuro-Markovian hybrids,” IVCNZ, pp. 320—326, 2003.

[24] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, “Deep learning for sensor-based activity recognition: A survey,” Pattern Recognit. Lett., 119, pp. 3-11, 2019.

[25] J. Yamashita, Y. Takimoto, H. Koya, H. Oishi, and T. Kumada, “Deep unsupervised activity visualization using head and eye movements,” IUI Companion, pp. 41—42, 2020.

[26] K. You, M. Long, Z. Cao, J. Wang, and M. I. Jordan, “Universal Domain Adaptation,” CVPR, pp. 2720—2729, 2019.