Vol.2022-EIP-95 No.1
2022/2/18

大量ウェアラブルデバイスと大規模生体情報時代における AI 機械学習のシャーロック問題への対策としての プライバシーエージェントの Docker ベクトル化

金子 格[1], 湯田 恵美[1]

概要：大量のウェアラブルデバイスにより高精度の生体情報が生成されるようになった．高精度で大量となった結果, プライバシー保護の観点で適切に処理することがより重要となった．一方 AI 機械学習の進歩により，大量の高精度情報から，より多くの個人属性を導き出し，個人の識別も可能になると予想される．それらによるプライバシー侵害が懸念されるが，それらから得られる情報の価値を考えると，その共有が抑制されれば社会的利益も大きく損なわれる．これをバランスさせる方法として，個人情報エージェントがすでに提案されているが，その具体的実装方法はあまり議論されていない．本報告ではプライバシーエージェントの Docker ベクトル化の可能性を論ずる． Docker ベクトル化によりプライバシーエージェントを標準的クラウド環境に実装でき, 生体情報などの個人情報を効率的に処理し，制御されたドメイン内での管理を維持することができる

キーワード：キーワード: 非均一ネットワーク，接触検出，感染者数, cocoa

Docker vectorization of privacy agents as a defense against the Sherlock problem of AI machine learning in the age of mass wearable devices and large-scale bio information.

Itaru Kaneko[1], Emi Yuda[1]

Abstract: Because of the large amount of highly accurate biometric information is generated by wearable devices, it became more difficult to handle these raw data appropriately in terms of privacy protection. Due to the development of AI machine learning in the future, more attribute of individual can be derived from large amount of high precision information, and it is also possible to identify individuals. Unauthorized use of them leads to infringement of individual rights but considering the value of the information obtained from them, it is a social loss to limit the use uniformly. A personal information agent has already been proposed as a method of balancing this, but the implementation method has not been discussed much. This report will describe Docker Vectorization of privacy agent. Docker Vectorization can be implemented in standardized cloud environment and may be executed efficiently and dynamically managed by the originator of personal information to maintain their usage within the controlled domain.

1. はじめに

本報告では，プライバシーエージェントの一実装法として, Docker ベクトル化を提案する．ウェアラブルデバイスは急速に普及しつつある．その多くがウエアラブルな生体センサであり，高精度で大量の生体情報，Massive Wearable Sensing Data 以下 MWSD)を取得できるようになった．AI 機械学習の進歩により, これらの生体情報のビッグデータから医療や公益に役立つ情報が得られると期待される．一方従来の予想を超える機微な情報が得られることがあり，プライバシー侵害や個人の権利の侵害を懸念する利用者も増えると予想される．データを共有し分析することから得られる情報の価値を考えれば，そのような懸念やリスクを最小化することにより，共有されるデ

u[1] 東北大学データ駆動科学・AI 教育研究センター, Center for Data-driven Science and Artificial Intelligence Tohoku University

ータを増大することが望ましい．

一方，中川は個人データの収集、管理、保護をおこなうパーソナル AI エージェントを提唱した[1]．データ主体の個人データとその利用条件をパーソナル AI エージェントによって管理することができる．従来の枠組みよりも個人情報の利用条件を柔軟に指定しうる魅力的な提案である．

一方エージェントは必ずしも AI である必要はないと報告者は考える．自ら選んだ代理人が設定した条件で，データが管理されることも一つの方法である．そして重要なのはエージェントの管理化で個人データをデータ処理する枠組みを実現することである．

本報告ではそのような個人データの情報処理を管理するエージェントをプライバシーエージェントと呼ぶ．そのような個人データの利用許諾を管理するプライバシーエージェントが満たすべき要求条件を検討し，その実現方法として Docker ベクトル化を提案する．

Docker ベクトル化の目的はプライバシーエージェントを標準的クラウド環境で効率的に実現することである．これにより生体情報などの個人情報を効率的に処理し，かつ，個人情報の分析方法を個人情報の提供者が管理し，利用方法と流通範囲を適切に制限することが可能であると考える．

2. MWSD と生体情報処理

2.1. MWSD の拡大

2020 年第 3 四半期の世界のウェアラブルデバイスの総出荷台数は 1 億 2503 万台，腕時計型は 3,290 万台，リストバンド型は 2,193 万台，耳装着型デバイスは 6,975 万台であった(図 1)．いずれも 10～50%の増加率で成長しており今後も急速な普及が進むと見込まれる．

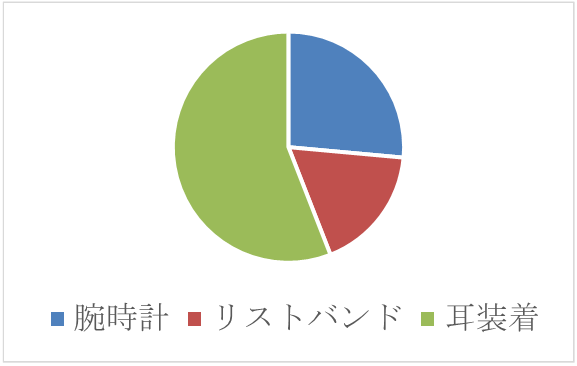

図 1 ウェアラブルデバイス出荷台数 1 億 2503 万台(2020 年) IDC の発表[2] による

これらの機器の主な用途は，電子決済や音楽再生，通知機能などであるが，デバイスの多くは心拍などの計測や加速度計も内蔵しており，これらから膨大な生体情報，MWSD が得られることになる．MWSD を利用することで体調管理や病気の予見などにも利用でき，その利用には大きな可能性が秘められている．筆者等は MWSD から様々な健康状態や心理状態が得られることを示した[3] [4]．

一方 AI 機械学習の進歩により，個人情報の提供者からみて想定を超える情報が抽出される（つまり，「そんなことまで知られてしまうのか」という感情を持たれる)ことへの対策が課題となりつつある．架空の探偵シャーロックホームズが，驚くべき推理で初対面の相手の職業や出身地を言い当ててしまうのに似ているので，筆者等はこれを「シャーロック問題」と呼んでいる．

MWSD の活用において，シャーロック問題は大きな障害となりえる．生体情報の提供をするデータ提供者は当然そのデータの利用目的に合意してデータを提供する．しかし，データの有効活用のために新たな分析手法も開発すべきだが，それがデータ提供者の想像を超える情報の暴露となりえる．結果として，データ提供に合意をすることにリスクを感じてしまう．またデータのより高度な利用が実現すればするほど，それまでデータ提供を行った利用者に不安をもたらす．

この問題の本質は，将来おこりうる最大の損失を見積もることが困難であることにある．したがってそのような不安を解消することが望ましい．

2.2. 著作権の支分権からの示唆

著作権の許諾にも似たような課題がある．いったん著作物の流通業者に著作権を許諾した後，許諾された著作物が予想を超えて利用されると，許諾の公正性に関しての摩擦が生じる．

しかし著作権においては比較的細かく具体的に支分権が設定されており，これが許諾時の想定が崩れる要素を減じている．そして想像しないような大規模な利用が発生して摩擦が生じる可能性を下げるのに寄与している．許諾を受ける側も認識の違いによる摩擦を避けるべきだから，あらかじめ可能性がある支分権は具体的に列挙して許諾を得ることに，著作権者との良好な関係を保てるというメリットがある．

プライバシー保護においても，利用許諾の範囲はできるだけ細かく具体的に示すことができれば，摩擦の解消に寄与すると考えられる．

3. プライバシーエージェント

3.1. エージェントによる同意形成

個人情報を利用する側はもちろんすべての利

用者から同一の許諾を得る必要はない．重要なのは異なる許諾ポリシーを判別して利用目的や処理方法に適合する個人情報を選択しセキュアな環境で処理できることであろう．

そこで考えられるのは，プライバシーエージェント，つまり許諾範囲を管理するソフトウエアに利用者の求める利用制限を代理させる方法である．著作権管理では広く使われていて，コピー保護，コピー回数制限など多様な制限を課す可能性があることは広く体験されている．個人情報にはプライバシーエージェントを付随させれば個人情報やその派生情報を常にデータ主体の管理下に置くことができる．

またプライバシーエージェントを更新可能であれば，あらたな利用条件を適応的に作ることが可能である．新しい分析手法が考案されそれにより新たな利用方法が生まれた場合にも，プライバシーエージェントによる制御であれば，そのような新たな利用方法の許諾をあらたに設け，利用者の許諾を求めればよい．

3.2. オブジェクト的プライバシー管理

次にこのようなプライバシーエージェントにより可能となるプライバシー管理を考えてみよう．

想定するのは個人情報のビッグデータにたいし AI 機械学習の手法によってなんらかの出力を得ようとする行為である．ここで問題となるのは，個人情報を提供した際に想定していなかったような分析手法について，どうすれば個人情報の提供者の適切な合意を得ながら利用できるか，という問題であった．

もともとどのような分析手法が可能となるかはわからないから，これを事前に合意をとることに困難である．

一方，新たな分析手法やアルゴリズムを示した上で許諾をとれば問題はない．

アルゴリズムによる許諾であれば，あるていどの分析手法の拡張も柔軟に対応可能である．そしてどの個人の個人情報をどの分析に提供するか，どこまでの分析を許すかが，フレキシブルにコントロールすることが可能だろう．

4. 個人データの Docker ベクトル化

4.1. ゼロトラストセキュリティモデルと多面的セキュリティモデル

このようなシステムで，プライバシーエージェントのセキュリティはどのように確保すべきだろうか．軽量でデータ主体からみて信頼できるセキュリティモデルであることが望ましい．

ゼロトラストセキュリティモデルは Marsh が 1994 年に単一のセキュリティ起点を持たないセキュリティのモデルとして発表した．筆者は 2001 年に，同様に単一のセキュリティ起点を持たない‘Multilateral Security Framework’を独自に提案した．単一セキュリティ起点を持たない点は共通であるが Marsh が進化シミュレーションによって実証的な分析を行ったのに対し，筆者はゲーム理論的にセキュリティ保持がナッシュ均衡を保つ条件を分析した．どちらの理論によっても分散セキュリティを実現可能であることが示され，今日単一セキュリティ起点を持たないゼロトラストセキュリティモデルは，標準的なセキュリティモデルとして普及しつつある．

プライバシーエージェントも利用者毎に複数のセキュリティエージェントを自由に選択，多重利用できるようにすれば，軽量で信頼性の高いセキュリティを実現できると考えられる．

4.2. プライバシーエージェントの要求条件

次にプライバシーエージェントが満たすべき要求条件を考える．プライバシーエージェントは複数セキュリティエージェントによる多重で柔軟なセキュリティエージェントによりセキュリティを実現可能であることが望ましい．

またデータ交換を行う複数のプライバシーエージェントが相互に認証可能であることが望ましい．

個人情報を処理する，個人情報利用者側のデータ処理モジュールは，これら複数のプライバシーエージェントから認証をうけた上で，許諾条件を確認できる仕組みがあることが望ましい．

たとえば個人データから何等かの情報を抽出するエンジンについてもプライバシーエージェントが抽出の可否を制御できるべきである．

データ処理の中間データや分析結果については，原則は平文ではなく暗号化してプライバシーエージェントの管理下におくことが安全だろうと考える．どのような派生データからもデータ主体の予想をこえた分析が可能にならないとは限らない．いつでもデータおよびそこから派生した分析結果の提供を止められることはデータ主体の不安を解消することに大きく貢献すると思われる．

それは派生データも個人情報と同じように保護することを意味するが，これも通常のコンテンツ保護技術に使われているコピー保護技術や，利用制限技術を用いて実現することが可能である．

4.3. クラウド実行環境

今日ほとんどのデータ処理は，クラウド環境上

で実行されている．以下では，環境上で，上記のような要求条件を満たす個人情報の処理方法を検討する．

第一にクラウド環境自体が信頼できるかという問題がある．しかしもしクラウド環境自体がそこで稼働するシステムの構造や設計に関係なく，プラットフォームとして情報漏洩を起こすものであれば，どんなシステムもクラウドプラットフォームを利用するかぎり，セキュアではないことになり，議論の意味がない．クラウドプラットフォームのセキュリティは絶対に必要なものとして前提として仮定することとする．

もちろんクラウドプラットフォームで稼働すればそれだけでセキュリティが保たれることは意味しない．クラウドプラットフォーム上には，無数の利用者が互いのセキュリティを確保しながら稼働しているが，問題はそれらがお互いのセキュリティ侵害や情報漏洩の可能性から自己の安全性を守れるかという問題である．

クラウドプラットフォームが保障するのは利用者間の独立性だけであり，クラウド上で稼働するVMは少なくとも独立したサーバと同様のセキュリティは持っているというところまでである．したがって本報告においては，そのような環境は前提として，データ主体が個人データの利用者や第三者に対しセキュリティを守れるかという点に絞って議論を進める．

4.4. 多面的セキュリティモデルによるComponentの信頼性評価

報告者は多面的セキュリティモデルによりシステムを構成する部品, Componentの信頼性を評価する方法を示した[8] [9] ．その要点を示す．

図 2 に多面的セキュリティモデルによるVerifierの相互認証を示す。

相互認証によりVerifierは相互にセキュリティ認証を行う．Verifier は Verification 性能により報酬を与えることによって Verification 性能を高めるインセンティブが与えられている．そしてこのような一定のインセンティブがあり，Verifierの互換性によってVerifierの交換コストがゼロであれば，ゲーム理論的上はVerifierが相互にセキュリティ不良の検出に全力を尽くすことがナッシュ均衡解となる．つまりVerifier群はお互いにセキュリティ不良の検出に全力を尽くすので，セキュリティ不良が見逃される可能性を小さくすることができる．

コンポーネントの多面的セキュリティモデルによる認証を図 3 に示す．

複数 Verifier から複数 Components を Verify する．すべての Verifier がすべての Component を Verify しなくても Verifier は他の Verifier の信頼性を評価できる．また各 Component はより多くの Verifier から Verification されることでより高い信頼性を得，このフレームワークによって各Componentは信頼性の指標を得ることができる．

筆者等はこのような枠組みと理論的分析の詳細を[8] [9] で論じている．

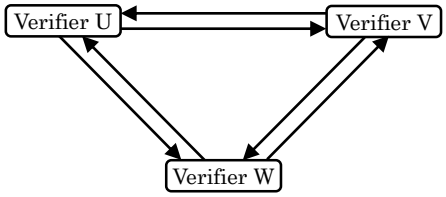

図 2 多面的セキュリティモデルの相互認証

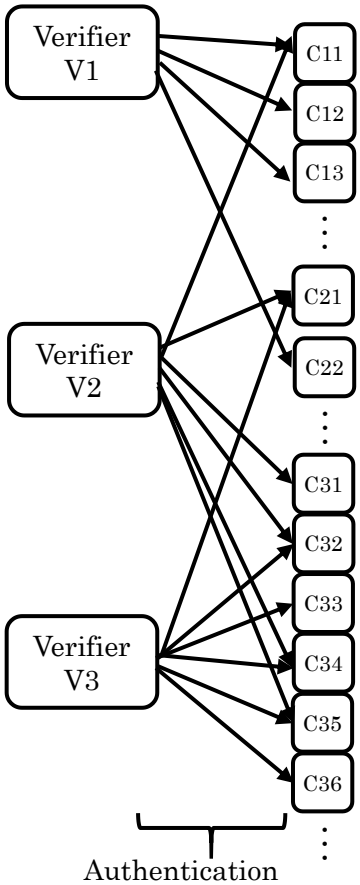

図 3 Component の多面的な認証

4.5. セキュリティエージェントの MSLM による実現

先に示した要件のために個人情報を提供した個人はそれぞれ自分が要求する個人情報の処理

上の制約下で自分の個人情報の処理が行われることを保証したい．そこでセキュリティエージェントを Component として実現してセキュリティエージェントが利用条件を確認してから処理を行うことで，利用者が求める個人情報の処理の制約を保証する．

具体的には利用者が求める個人情報保護上の基準に適合していることを確認するモジュールを，そのデータの処理を行う条件として設定する．

これによって，個人情報取り扱い事業者が実行するデータ処理は，個人情報の元の提供者が要求する監査基準を満たしたものに限られることが保障される．

Verifier 同様に，個人情報保護を目的とした監査モジュールを複数指定することも当然可能である．このように複数の監査機関を要求することを可能とすることで，「多重監査」や「相互監査」が可能となるので，一つの監査機関によらない分散監査によってシステムの安全性をより信頼できるものとすることが期待される．

4.6. Docker Container 化

個人情報エージェントは図 4 に示すよう Docker Host 下のコンテナとして実行するのが可能だと考えられる．

クラウド環境では Verifier や Component はまずイメージとして作成され，Docker Container 中に含めて実行することが考えられる．Verifier と Component がイメージとして提供され各個人の要望に応じてコンテナ化されていれば，各個人の個人情報に関する要求事項にあわせた処理を行うことが可能であろう．

この方法はあらゆる処理や認証方法を個人情報に結び付けることが可能であるから，非常にフレキシブルで強力である．

また Docker Container はそれほど大きくなく(数 100Mb におさまり)，Docker Host は Docker Container に含まれているイメージを実際にいつもロードするわけではなく，同じイメージがすでにメモリ中に存在すればそれを共用することが可能であるため，同じイメージの集合から構成あされる膨大な Docker Container を起動・実行してもそのメモリコストや計算機コストはごくわずかに抑えることが可能である．

したがって，個々の個人情報に Docker Container が添付されている，という構成は構造をよくしらなければやや冗長に見えるが，実際にはきわめて効率的に実行することが可能であると考えられる．

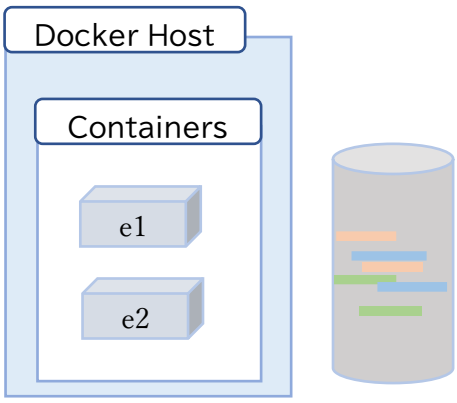

図 4 Docker Container 化されたプライバシーエージェント

現在 ISO/IEC 23092 シリーズで DNA のゲノム情報の符号化の標準化作業が行われており，将来ゲノム情報のデータベースが医療や健康産業目的で大量に国際間でもやりとりされる場合にその適切な管理方法が課題となっている．その場合はこの個人情報の保護手段を Docker Container 化して個々のゲノム情報に付随させて伝送保存するという方法が効率的かつ効果的であると考えられる．

4.7. Docker Vector 化

とはいえ Docker Container は必ずしも万能とはいえない．Docker Container のサイズはイメージがコンパクトになったとはいえ数 100Mb に及ぶ場合がある．最低でも数 GB のゲノム情報本体の記述データとしては数 100Mb の Docker Container のオーバーヘッドは許容範囲だろう．しかしウエアラブル機器から得られるデータは，たとえば毎時間の体温といった場合たかだか 1kb 程度のことがある．1kb のデータ本体に対し数 100MB のオーバーヘッドは大きすぎる．そこで，Docker Vector 化することを考える．

Docker Vector 化を図 5 に示す．Docker Container をそのまま扱うのではなく，Docker Container の記述フォーマットに基づくディスクリプタ PPPD(Privacy Protection Policy Descriptor)で記述する．PPPD にはプライバシーエージェントを実現する Docker Container 自体は含めず，あらかじめ登録した Docker Container の識別子と個人情報管理のパラメータを含める．PPPD は json などで記述する．

プライバシーエージェントを実現する Docker Container を識別子で表すことで PPPD の符号化効率は大幅に向上する．一方 128bit 程度の bit 幅があれば，全人類に対し毎秒 1 種類の新しいプライバシーエージェントをあらたに定義しても符号空間はまったく枯渇する心配はない．

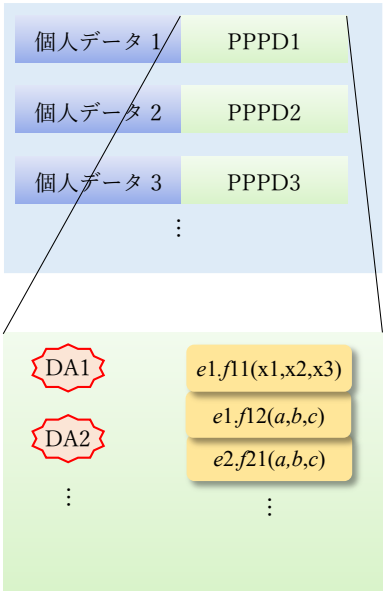

図 5 Docker Vector 化

5. まとめ

本報告ではウエアラブル機器から得られる大量の生体情報，MWSD の個人情報保護に有効な個人情報保護の技術的手段として，Docker Vector 化を提案した．ウエアラブル機器から得られる大量の生体情報は今後蓄積とあらたな分析手段によって現在予想しないような様々な機微な情報がそこから抽出することが可能となると考えられる．したがって生体情報の提供が合理的な基準で行われるためには，その保護に柔軟で強力な方法が利用できることが望ましい．それには相互認証や多重評価が可能でアルゴリズムによって個人情報の保護水準や派生した情報の管理が可能である方法が望ましい．そのようなモジュールは Docker Image から構成した Docker Container として記述，提供すればゲノム情報処理などでは効率も十分で個人情報保護機能は十分柔軟で強力なものとすることが可能であると思われる．しかしウエアラブル機器の生体信号の保護に Docker Container ファイルをそのまま付随させる方法は，オーバーヘッドが大きすぎる．そこで本報告では Docker Container そのものではなくその記述子，PPPD によって個人情報保護の技術的手段を記述することを提案した．

この仕組みのうち Docker Container は標準の仕組みを用いるから特に新たに実装することは必要ない．必要なのは PPPD であるがこれは json ファイルとして比較的簡単に記述することが可能であると考える．

Docker Vector 化は単に登録済の json ファイルを参照する仕組みである．そこで検証すべきは個人情報データを Docker Container で効率的に処理可能かという問題になる．この点については複数の個人情報に対し何種類かのプライバシーエージェントが存在しそれに対して分析を行う場合に，容易に記述が可能であるかを検証する必要があり，また当然この目的のための標準的な json 記述子を検討する必要がある．これらの検討が今後の課題であると考える．

参考文献:

[1] 中川 裕志，AI 倫理指針の動向とパーソナル AI エージェント，情報通信政策研究/3 巻 (2019) 2 号 (2019)
[2] IDC, 2020 年第 3 四半期 世界／国内ウェアラブルデバイス市場規模を発表, https://www.idc.com/getdoc.jsp?containerId=prJPJ47221920(2021/1 取得)
[3] Junichiro Hayano, Tetsuya Tanabiki, Shinichiro Iwata, Katsumi Abe, Emi Yuda. Estimation of Office Worker's Emotion Types Using Two-dimensional Model Consisted of Biometric Signals, INTERNATIONAL JOURNAL OF AFFECTIVE ENGINEERING 20(2) 105-110 (2021)
[4] Itaru Kaneko, Yutaka Yoshida, Emi Yuda, Junichiro Hayano. Sensing of Microvascular Vasomotion Using Consumer Camera. Sensors (Basel, Switzerland) 21(18) 6256-6256 (2021)
[5] 金子格，湯田恵美，吉田豊，ホルター心電計の内蔵加速度センサを用いた活動推定の改良と個人情報保護に関する考察，情報処理学会研究報告
[6] Stephen Paul Marsh,Formalising Trust as a Computational Concept, Computing Science and Mathematics eTheses, http://hdl.handle.net/1893/2010 (1994)
[7] L. Mui,M. Mohtashemi,A. Halberstadt,A computational model of trust and reputation, Proceedings of the 35th Annual Hawaii International Conference on System Sciences(2002)
[8] I. Kaneko; K. Shirai, The multi-lateral security framework for the ubiquitous audiovisual services, 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace
[9] 金子格，確率的多面的セキュリティモデルとブロックチェーンを用いたメディア IoT 向け軽量セキュリティフレームワーク，電気学会論文誌 C（電子・情報・システム部門誌）/137 巻 (2017) 6 号