

# 仮想人体生成モデルにおける品質管理

尾藤 宏達 日比 壮信 大野 健太 齊藤 友樹 南 和宏 丸山 宏

仮想人体生成モデルは人体に関する 2,000 種以上の様々な属性の関係を同時分布として表現したものであり、広い意味で生成モデルである。特定の利用目的に特化したものではなく、API を通して様々な利用方法を想定しているため、通常の機械学習システムとは異なる品質管理の方法が必要となる。本稿では、仮想人体生成モデルの品質に関して、1) 品質マネジメント、2) 訓練データの品質管理、3) モデルの品質管理、4) プライバシーの評価、5) その他の品質についての考慮点について、我々がどのような設計を行ったかについて述べる。

## 1 はじめに

人々がより健康で豊かな生活を送れるようにする様々な技術と、それらを組み合わせたソリューションの実現に高い関心が寄せられてきた。これらの技術開発やソリューションの実現に有効なツールと期待し、2022 年 2 月に花王株式会社 (以下、花王) と株式会社 Preferred Networks (以下、PFN) は、ヒトの多様な身体の状態や生活習慣などで記述されたデータに基づいて構築した汎用の統計モデル、仮想人体生成モデルを開発し、2023 年 2 月より商用の API サービスとして提供を開始した。

特定目的を持つ機械学習モデルの品質管理については、いくつかのガイドライン [2] [15] が知られているが、生成モデルのように目的を限定しない、汎用の機

械学習モデルの品質管理ではカバーできない考慮点がある。また、API を通して商用のサービスを行うには、ガバナンスを含めた品質マネジメントを包括的に考える必要がある。

本稿では、仮想人体生成モデルの概要を 2 章で紹介した後、仮想人体生成モデルの品質マネジメントの考え方について 3 章で述べる。その後、訓練データの品質管理について 4 章で、モデルの品質管理について 5 章で、プライバシーについて 6 章でそれぞれ解説する。本稿でカバーできなかったその他の考慮点については、7 章で議論する。

## 2 仮想人体生成モデルの概要

仮想人体生成モデルは、人体の様々な観測可能な計測値 (以下、属性と呼ぶ) の出現パターンを同時確率分布の形で表現したものである。属性としては、身長や体重などの身体計測値、血圧や脈拍などのバイタルサイン、血液検査、疾病の有無などに加え、生活習慣、運動機能の状態、認知機能の状態、皮膚や毛髪の状態、性格特性、ストレスや睡眠の状況、食習慣などを尋ねる質問票による項目で構成されている。これらの属性は、採血を伴う侵襲的な計測から質問票に基づいた計測で得られるものまで含まれており、検査にかかるコストも多様である。2023 年 7 月現在、最も多くの属性数を持つモデルでは、2,150 属性の同時確

Quality Management for Virtual Human Generative Model

Kotatsu Bito, Yuki Saito, 花王株式会社 デジタル事業創造部, Digital Business Creation, Kao Corporation.

Masanobu Hibi, 花王株式会社 生物科学研究所, Biological Science Research Laboratories, Kao Corporation.

Kenta Oono, 株式会社 Preferred Networks, Preferred Networks, Inc..

Kazuhiro Minami, 統計数理研究所, The Institute of Statistical Mathematics.

Maruyama Hiroshi, 花王株式会社, Kao Corporation.

率分布となっている。この仮想人体生成モデルは、主に日本の成人男女のデータによって訓練されていて、日本の成人男女がどのような属性値をとるかの統計的なパターンを記憶していると考えられる。ある条件の下での条件付き分布を求めることで、入力した条件に基づいて条件付き分布を生成する、広い意味での生成モデルとして捉えることができる。このモデルは特定の利用目的に特化したものではなく、ヘルスケアアプリケーションでの活用を含め多様な利用方法が期待された。そのため、APIを介して提供することが好ましいと考え、商用のAPIサービス(VITA NAVI)として、提供を開始した。

### 3 仮想人体生成モデルにおける品質マネジメントの考え方

品質マネジメントの世界標準 ISO9000 によれば、品質とは「製品やサービスなどに本来備わっている特性の集まりが、顧客の要求事項を満たす程度」とされている。我々が直接想定する顧客は、APIを利用して例えば健康関連のアプリケーションを作り運営する事業者だが、その先にはアプリケーションを利用する一般の利用者がいるはずである。本稿では、下記4種のステークホルダーを広く顧客と考え、これらのステークホルダーが期待する特性を実現することが、仮想人体生成モデルの品質と捉える。

- **事業者** 仮想人体生成モデルのAPIを利用してアプリケーションを作り運営する事業者。仮想人体生成モデルが提供する機能がビジネス目的に合致し、必要な性能で動くことを期待する。
- **エンドユーザー** 上記アプリケーションを使う個人。一義的な期待は事業者が提供するアプリケーションにあるが、その裏にある仮想人体生成モデルが、自身の健康や Wellbeing などにポジティブに寄与することを期待する。また、自身のプライバシーが適切に守られることを期待する。
- **データセット提供者** 被験者を測定し、訓練データセットを生成する者(4章参照)。訓練データセットの権利が適切に守られることを期待する。
- **被験者** 上記訓練データセットを作成するために測定される個人。測定にあたり自分に不利益が

生じないことを期待する。

これらの期待に対して、仮想人体生成モデルの品質マネジメントを、以下の項目に従って設計した。

1. モデルの継続的な改善
2. 顧客の期待値マネジメント
3. コンプライアンス
4. 透明性
5. ガバナンスとプロセス

それぞれについて、以下の節で順に述べる。

#### 3.1 モデルの継続的な改善

仮想人体生成モデルは新しい考え方に基づいて作られた統計モデルであり、訓練の方法や評価指標も、顧客のフィードバックを得ながら探索的に決めていく必要があった。このため、我々はまず、継続的な改善を繰り返していくサービスとして仮想人体生成モデルを設計した。予め決められたスケジュール(現在、ほぼ毎月)で、我々は新しいモデルを発行する。新しいモデルには、以下のような変更を伴う。

1. スキーマの変更。新しい属性が加わる、使われない属性を削除する、属性のデータ型(数値とカテゴリ値)の定義を変更する、など。
2. 訓練データの前処理。訓練データの外れ値処理に改善が必要な場合、など。
3. モデルの改善。訓練アルゴリズムやハイパーパラメタの変更によるもの。

事業者のアプリケーション継続性を担保するため、一度発行したモデルは最低6ヶ月間維持され、モデルのサービス終了の場合も十分なリードタイムを持って事業者へ通知される。また、軽微なバグフィックスを除けば、一度発行したモデルはそのサービス終了まで変更しない。その間、同じモデルに対して同じ入力を与えれば、常に同じ出力が得られることを保証する。APIの振る舞いをこのように固定することにより、事業者のアプリケーションの開発・運用の負担を軽減することを狙った。

#### 3.2 顧客の期待値マネジメント

機械学習あるいは生成モデルなどの技術を使ったシステムは「AI」と称されることがある。しかし、「AI」

は聞く者によって解釈の異なることがあり、誤った期待を与えやすい。仮想人体生成モデルは訓練データにおける属性間の同時分布を近似した統計モデルであり、それ以上でもそれ以下でもない。特に、2つの属性の間に、何らかの形の関係があった場合に、そのことはそれらの値が訓練データセット中に同時に観測されることが多いと推定されたこと示すのみであり、それらの間の因果関係について何らの示唆をするものではない。

仮想人体生成モデルの1つの利用シナリオとして、「もしこうだったら」という反事実の推定を行うことができる。例えば、「私が毎日運動していたとしたら、私の内臓脂肪面積はいくらと推定されただろうか」というクエリを出すこともできる(我々はこれを**what-if 分析**と呼んでいる)。これは、自分の健康状態を考える際に、可能なシナリオを検討するのに使えるが、それを因果関係と解釈されないよう、仮想人体生成モデルのガイドブックなどに丁寧に解説することによって誤解を回避しようとしている。

仮想人体生成モデルを使ってある属性の値を推定したとき、その値が事業者やエンドユーザーの期待と異なることは十分にあり得る。それは1) 訓練データの分布がそのようになっていたか、さもなくば2) 訓練における誤差により、訓練データの分布から外れた値が返されたか、のいずれかである(モデル近似誤差の評価指標については5章で述べる)。仮想人体生成モデルが訓練データセットの同時分布を近似する統計モデルに過ぎない、ということをあらゆる機会を捉えて明示することにより、ステークホルダの理解を高める努力をしている。

### 3.3 コンプライアンス

他のドメインの機械学習モデルと異なり、仮想人体生成モデルは個人の健康状態に関する情報を扱う。このため、医薬品、医療機器などの品質、有効性及び安全性の確保等に関する法律(薬機法)及び個人情報の保護に関する法律(個人情報保護法)を遵守する必要がある。エンドユーザーの保護のため、我々は仮想人体生成モデルがプログラム医療機器でないことを確認し、また事業者にも、仮想人体生成モデルの出力

だけに基づいて医療行為を行うことを禁じている(詳細はAPIサービスの利用規約に記されている)。

エンドユーザーの自己の健康状態など要配慮個人情報(API経由でやり取りされることが考えられることから、仮想人体生成モデルのAPI利用にあたっては、その入出力データを保持せず、API呼び出しが終わったら廃棄するようにしている。それらのデータを将来の訓練に用いることも行わない(これには訓練データの品質を維持するため、という理由もある)。

訓練データ収集の際の被験者の保護のためには、文部科学省、厚生労働省及び経済産業省による「人を対象とする生命科学・医学系研究に関する倫理指針」に基づき、花王及びPFNの倫理審査委員会の審査を経て収集したデータ、あるいは個人情報保護法に決められた匿名加工情報取扱事業者から購入した匿名加工情報のみを扱う。さらに、プライバシーに関しては様々な攻撃手法が考えられることから、最新のプライバシー手法での評価も行う(6章参照)。

### 3.4 透明性

主たる訓練データを取得するための試験プロトコールと訓練アルゴリズムの設計については、透明性の確保のため、訓練データセットの設計及び訓練アルゴリズムの詳細について、それぞれの論文を公開し、ステークホルダがその内容を精査できるようにしている[6][11]。

### 3.5 ガバナンスとプロセス

上記すべての観点について確実に品質改善を行っていくために、仮想人体生成モデルの品質管理全体に関わる意思決定機関として、モデル開発及び利用に関わる複数のステークホルダからのメンバーで構成される**スキーマ委員会**を置く。スキーマ委員会は月1回開催され、翌々月に発行されるモデルの詳細を決定する。この決定に基づき、訓練データセットを準備し、訓練を行い、5章に述べる品質指標に基づいて評価を行い、モデルを発行する(図1)。

仮想人体生成モデルのAPIサービスとしての開発・運用は、スクラムによるDevOps(あるいはMLOps)で行っている。図1に示すスキーマ委員会の決定に

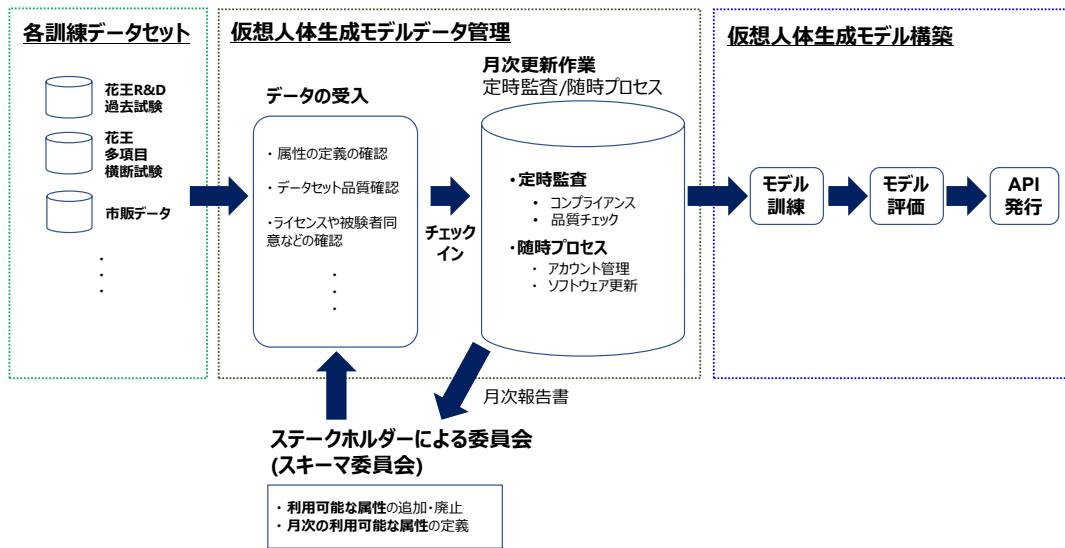


図 1 訓練データセットの取得から API としてのモデル発行までのプロセス。

基づいた毎月の訓練データセットの点検と更新，提供された訓練データセットに基づいたモデルの訓練，評価及び API を通したリリース・運用を行う。また，API のリリースと連動する形で，API 利用を支援するツール（モデルの各属性に一意に紐づく符号と対応する属性の名称の変換，各属性の定義や説明などの情報を提供するツール）などの周辺サービスの拡充と更新も行っている。この様なモデルの発行サイクルを起点とした運用を回していく形で，事業者の期待に機敏に応えられる体制を整えている。

#### 4 訓練データの品質管理

仮想人体生成モデルでは，2,000 属性を超える非常に多くの属性を扱う。単一の試験で，全ての属性を計測を同時に行うことは現実的に困難であるため，異なるデータソースからなる複数のデータセットから構築することを検討した。これらのデータセットは，適切にデザインされた試験によって得られ，品質が十分であるものを利用すべきであると考えた。そこで，私たちは，レコード数と属性数の異なる次の 3 種類からなる階層的なデータセットを利用した。

1. 属性数が多く（数百～数万），被験者数の少ない（数十～数百），花王の研究開発部門が過去に実施

した試験のデータ。

2. 属性数が中程度（千以上）で，被験者数が中程度（数百以上）である，花王が後述の方法で新たに実施した試験のデータ。
3. 属性数が少なく（数百以下），被験者数の多い（数十万以上），市販の健康診断結果とレセプト（診療報酬明細書）の情報に基づくデータ。

本章では，上述のデータセットの 1 つである，花王が実施した多くの検査項目からなる横断的試験（多項目横断試験）[6] のデータ取得を例として述べる。

仮想人体生成モデルでは，利用する訓練データは主に観察研究のうちでも横断的試験と呼ばれる研究デザインで得られたものを中心に，ある時点での人体の属性の出現パターンのモデル化を行っている。近年，観察研究の評価の一般化を可能にするべく，観察研究の正確かつ完全なる報告に含まれるべき事項に関する推奨として定められた STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) 声明[14] が定められている。本試験は，その STROBE 声明に基づき試験デザインが設計されていることを示し，加えて試験の透明性を確保するために，試験デザインをプロトコル論文としてまとめ，査読付き論文誌に投稿している [6]。本試験は，仮想

人体生成モデルを構築するデータセットの取得を目的として、日本の年齢分布に合わせた成人男女 1,000 名を目標として、試験を実施した。サンプルサイズは、予備的な精度の見積もりに基づいて決定した。データには、多様な属性で人体を記述できる様に、身体計測、血液、尿、唾液、経口ブドウ糖負荷試験から血液の生化学項目、糞便、顔面皮膚、頭皮皮膚、唾液からの菌叢プロファイル、顔面及び頭皮表面脂質の mRNA (SSL-RNA [7])、プロテオーム、代謝物分析、生活習慣調査及びアンケート、認知機能検査、運動機能検査、血管機能検査、脱毛症検査及び体臭検査などが含まれている。各被験者のデータは、電子データベースと症例記録用紙を用いて管理された。今回、2つのレベルのデータの品質管理を行った。第 1 レベルの品質管理を、試験の調査担当者が症例記録用紙に記録を入力する際に行った。第 2 のレベルの品質管理として、独立したグループがデータの監視と検証を行った。電子データベースは、正確なデータ入力であるかをさらに確実にするための、二重入力システムとして機能した。この様に、仮想人体生成モデルの構築に供される訓練データは、データ取得のためのデザインの妥当性やその後のデータマネジメントを行うことで品質の担保を行っている。

## 5 モデルの訓練

本章では、仮想人体生成モデルのアルゴリズム及び訓練方法の概要と、その評価方法について述べる。

### 5.1 モデルのアルゴリズム及び訓練方法の概要

仮想人体生成モデルで用いる属性は、数値とカテゴリ値の両方を含んでいるため、異なるデータ型を扱うことができるモデルを選択する必要がある。そのため、Heterogeneous-Incomplete Variational Autoencoder (HIVAE) [10] を利用した。HIVAE は変分オートエンコーダ (VAE) [8] の拡張であり、属性のタイプに応じて異なる尤度関数を選択できる。例えば、体重と身長には正規分布を、生物学的性別 (男性、女性) にはカテゴリ分布を入力として利用できる。観察された属性のセットなどを入力として、訓練されたモデルを使用し、各属性の (条件付き) 確率分布のパラ

メータを、入力に用いなかった属性を含め推定できる。我々は HIVAE の学習に、近年の言語モデルで事前学習に採用されている、マスク言語モデル [5] と類似した手法を採用した。通常の HIVAE は、Evidence Lower Bound (ELBO) と呼ばれる、モデルが入力をどのくらい復元できるかを表す指標を最大化するように学習される。それに対し、我々は、入力をランダムに欠損させ (マスクを施し)、欠損した入力からマスクした値を予測するようにモデルを学習した。学習の目的関数をこのように変更することで、欠損補完の精度が向上することが数値的に確かめられた。

この際、仮想人体生成モデルの構築では、花王で品質検査を行った後に、花王での評価に用いるための一部のデータを除き (ホールドアウトセット)、PFN にモデル構築のためのデータセットとして提供され、モデルの構築が行われている (図 1)。このアルゴリズムの詳細についても、透明性を高めるため、論文の形式にまとめプレプリントサーバにおいて公開している [11]。

### 5.2 モデルの評価指標

特定の利用目的に用いる機械学習モデルでは、モデルの評価はホールドアウトセットによる精度で行うのが普通である。しかし、仮想人体生成モデルにおいては、1) 一般に推定値は確率分布 (条件付き分布) の形で得られること、また 2) 2,000 を超える属性の任意の組み合わせが、入力・出力になりうることで、それらの精度を網羅的に評価するのは現実的でないこと、の 2 点を考え、評価指標を設計した。この評価指標は、1) 欠損属性の予測精度指標と、2) 経験分布 (訓練データ) の統計的性質をどの程度再現できているか、という 2 つの異なる視点からなる。

予測精度の視点では、ホールドアウトセットに対していくつかの入出力のパターンを固定し、実値とモデルの推定値 (確率分布の最頻値) と比較して予測精度を測定した。評価指標を決めるにあたって、事実上無限にある入出力パターンのうち、どれを選ぶかは非常に重要である。多くの組み合わせにおいては、属性間に関連がなく、事実上独立 (あるいは非常に弱い関連がノイズに埋もれている) とみなせる。我々は、想定

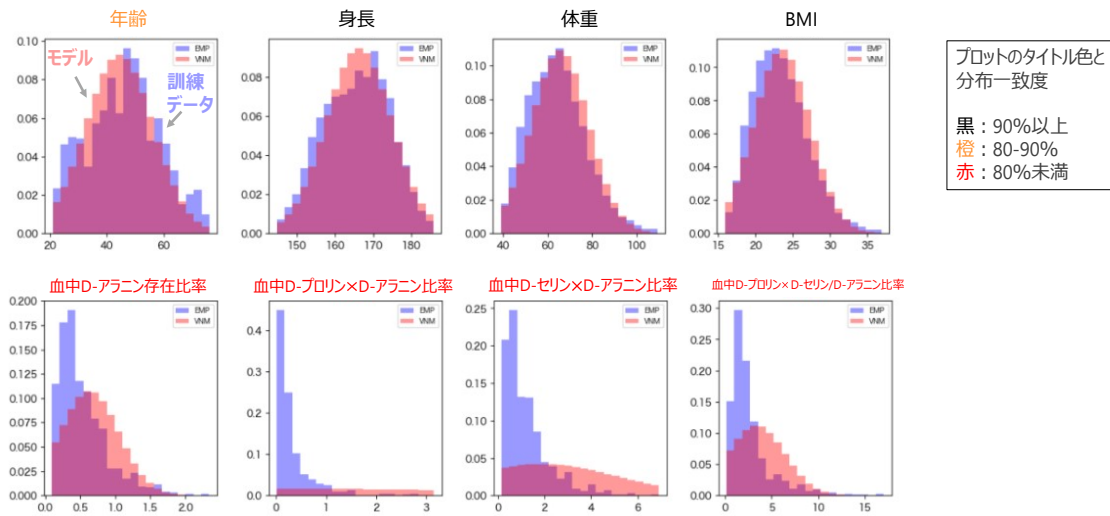


図2 訓練データとモデル出力の周辺分布の比較: 各分布の面積は1に規格化しており、縦軸は確率を示している。横軸は各属性の値である。分布の一致度は各分布の重なり面積の割合で定義した。上段に分布の一致度が高いものの例として4属性を、下段に分布の一致度が低いものの例として4属性を示す。

される典型的な利用シナリオを考慮しながら、属性間に一定以上の関連が見られるものをいくつか評価用パターンとして抽出し、それらの予測精度を全体の指標とした。なお、この評価用パターンは、モデル開発者には知らされていない。

仮想人体生成モデルは同時確率分布であり、訓練データの分布(経験分布)をよく反映していることが期待される。このため経験分布の再現という第2の視点では、訓練データの様々な統計的性質が、モデルの統計的性質とどの程度一致するかを評価した。同時分布の統計的性質のうち、それぞれの属性の周辺分布が一番わかりやすいもので、これについてはモデルの出力と訓練データの統計量を、すべての属性について比較する。我々はこれを **unary** 指標と呼ぶ。図2に、訓練データとモデル出力の周辺分布の比較を示す。上段の例の様に、多くの属性で一致度が高いことが認められる。一方で、モデル訓練において仮定している分布が適していないものも一部認められており、見直しを進めている。

一方、多くの場合モデルはある属性集合を入力として、他の属性の条件付き分布を求める、という形で使

われるので、複数の属性間の関連を調べなければならない。その場合、予測精度の評価と同様、入出力のパターンの組合せ爆発の問題が起きる。我々は、強い関連性を持つと考えられるいくつかの属性ペア(例:空腹時血糖とHbA1c)について、モデルと訓練データのそれぞれの条件付き分布を比較することで、両者に顕著な相違がないことを確認している(**binary** 指標と呼ぶ)。図3に、22年12月にリリースしたモデル(fomalhaut)と23年5月にリリースしたモデル(intercrus)それぞれの、4つのペアのbinary指標を示す。モデルの訓練アルゴリズムの改善活動によって、新しいモデルで訓練データとモデルの出力する同時分布との一致度が向上していることがわかる。

3章で述べた定期的なモデルのアップデートを実施する度に、予測精度と経験分布の再現という、2つの視点で評価を行い、全体として品質が改善されていることを確認している。

## 6 プライバシーの評価

仮想人体生成モデルで生成されたデータは、合成データであり、匿名加工処理されたデータとは異なる

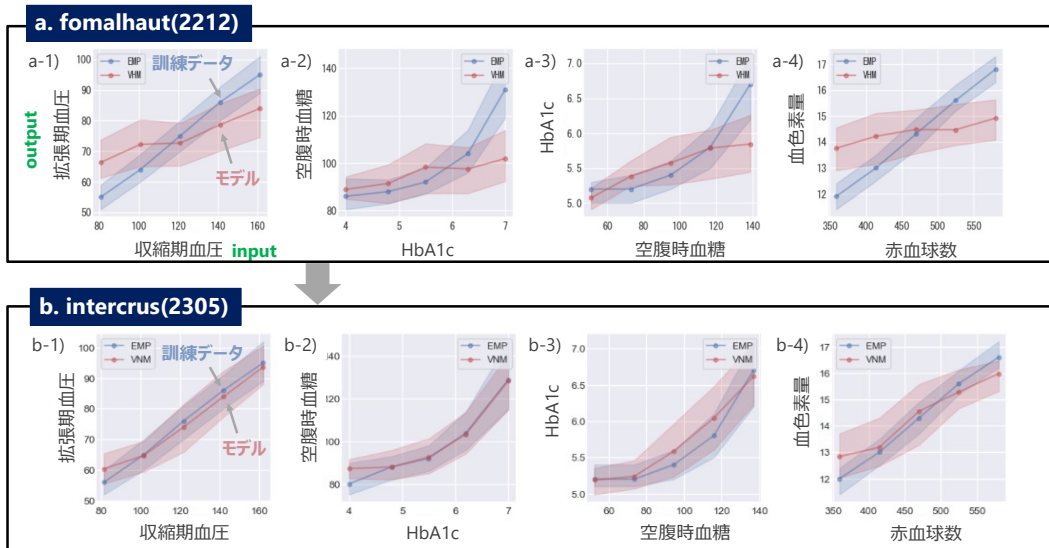


図 3 2つモデル, (a)fomalhaut と (b)intercrus の主要な属性ペアの関係 (条件付き分布) の比較: 訓練データでは, 横軸に用いた属性の平均から標準偏差の 3 倍以上離れた値を除き, 5 等分した各範囲の中央の値に対して, 各範囲における縦軸の属性の中央値をプロットし作成した. 縦軸の帯は該当属性の第一四分位と第三四分位の範囲を示している. モデルでは, 訓練データより得た 5 つの範囲の中央の値と, 各中央の値を条件とした確率分布に基づき生成した各 100 レコードの縦軸の属性の中央値をプロットし作成した. 縦軸の帯は生成したレコードの縦軸の属性の第一四分位と第三四分位の範囲を示している.

り, レコード識別や性別, 年齢, 職業といったレコードのデモグラフィック情報の属性推定といった問題は起きないと考えられる. しかし, 既存研究でいくつかの漏洩シナリオが指摘されている [1][4][9][12]. 特に極端な値の情報が精緻な学習モデルから漏洩するリスクがあることが指摘されている [13]. この安全性の評価指標として, 元データとの類似性やメンバーシップ推定の成功率が考えられる. 今回は, 元データの類似性 Distance to the closest record (DCR) の指標を用いる方法を検討した. 生成したデータから抽出した外れレコードと訓練データのレコードの DCR を計測し, 外れレコードの DCR が全体のレコードの DCR の平均よりも際立って小さいといったことがないかの確認を行った. 仮想人体生成モデルで 1,000 レコード生成し, k-means 法によるクラスタリングを行って, 単独のレコードでクラスタを形成するものを外れレコードとした [3]. 検討の結果, 5 つ抽出された外れレコードの DCR は, 全体のレコードの DCR の

平均よりも大きいことが確かめられた. よって, 本評価においては, 訓練データに含まれる外れレコードを現時点の仮想人体生成モデルで再現されるリスクは確認できなかった. これは, 本モデルが十分に汎化されているからと考えられた. ただし, 今回は単独の外れレコードを対象にしたが, 少数人で形成されるクラスタのケースを確認できていないのと, DCR が機密性に関するセマンティックスを適切に捉えていない可能性も考えられた. 今後, モデルの推論性能の改善に併せて, 別の観点の評価も加えた多面的なプライバシーリスクの評価を行っていく必要がある.

## 7 その他の品質に関する考慮点

本稿では詳細は述べないが, 以下にその他の品質に関する考慮点について列挙する.

- システムレベルの品質 本稿のスコップ外だが, CI や構成管理など, 通常のソフトウェア開発・運用におけるベストプラクティスを適用している.



- **インシデント・レスポンス及びBCP(事業継続計画)** 現在のところ、アドホックな対応だが、今後利用が拡大するにつれ、導入していく予定である。
- **カスタマー・サービス** 現在のところ、営業担当者による個別対応を行っている。今後、ユーザー・コミュニティ形成などを通して、スケールできるサービスを提供していきたい。
- **モデルの拡張** 特に、新しいデータセット(異なるデモグラフィック、新たな属性など)を導入する場合、ステークホルダー間の利害関係が対立する場面も考えられる。

昨今、特に言語、画像、音声などの領域における生成モデルの発展が著しい。人体のデータによる生成モデルも今後目覚ましく発展し、様々な技術やサービスが生まれるものと期待される。一方で、人体のデータの取得や利用には、倫理、法規制などを十分に配慮する必要がある。人体データの生成モデルを社会実装する上で、人体のデータを利用するが故に考慮すべきことがあり、データの取得、モデルの構築、モデルの提供までの一連のプロセスにおける品質管理について、引き続き十分に検討されていく必要があると考えられる。また、法規制などを含め、サービスを取り巻く環境も目まぐるしく変化していくことから、変化に対応可能な形でのサービス開発の体制と運用も重要である。今後、実践の中から技術開発と社会実装の双方のバランスを考慮した、より良い方法論が生まれていくことが期待される。

**謝辞** 本稿に示された内容は、花王とPFNの担当者と共に進めた検討の一部を、日本ソフトウェア科学会機械学習研究会(MLSE)夏合宿2023で発表した内容に基づき再構成し、論文としてまとめたものである。各プロジェクトメンバーに、深く感謝する。

## 参考文献

- [1] Abowd, J. M. and Vilhuber, L.: How protective are synthetic data?, *International Conference on Privacy in Statistical Databases*, Springer, 2008, pp. 239–246.
- [2] AIプロダクト品質保証コンソーシアム(QA4AIコンソーシアム): *AIプロダクト品質保証ガイドライン* 2023.06版, 2023.
- [3] Chandola, V., Banerjee, A., and Kumar, V.: Anomaly detection: A survey, *ACM computing surveys (CSUR)*, Vol. 41, No. 3(2009), pp. 1–58.
- [4] Charest, A.-S.: How can we analyze differentially-private synthetic datasets?, *Journal of Privacy and Confidentiality*, Vol. 2, No. 2(2011).
- [5] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota, Association for Computational Linguistics, June 2019, pp. 4171–4186.
- [6] Hibi, M., Katada, S., Kawakami, A., Bito, K., Ohtsuka, M., Sugitani, K., Muliandi, A., Yamanaka, N., Hasumura, T., Ando, Y., et al.: Assessment of Multidimensional Health Care Parameters Among Adults in Japan for Developing a Virtual Human Generative Model: Protocol for a Cross-sectional Study, *JMIR Research Protocols*, Vol. 12, No. 1(2023), pp. e47024.
- [7] Inoue, T., Kuwano, T., Uehara, Y., Yano, M., Oya, N., Takada, N., Tanaka, S., Ueda, Y., Hachiya, A., Takahashi, Y., et al.: Non-invasive human skin transcriptome analysis using mRNA in skin surface lipids, *Communications Biology*, Vol. 5, No. 1(2022), pp. 215.
- [8] Kingma, D. P. and Welling, M.: Auto-Encoding Variational Bayes, *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.
- [9] McClure, D. and Reiter, J. P.: Differential Privacy and Statistical Disclosure Risk Measures: An Investigation with Binary Synthetic Data., *Trans. Data Priv.*, Vol. 5, No. 3(2012), pp. 535–552.
- [10] Nazábal, A., Olmos, P. M., Ghahramani, Z., and Valera, I.: Handling incomplete heterogeneous data using VAEs, *Pattern Recognition*, Vol. 107(2020), pp. 107501.
- [11] Oono, K., Charoenphakdee, N., Bito, K., Gao, Z., Ota, Y., Yamaguchi, S., Sugawara, Y., Maeda, S.-i., Miyoshi, K., Saito, Y., et al.: Virtual Human Generative Model: Masked Modeling Approach for Learning Human Characteristics, *arXiv preprint arXiv:2306.10656*, (2023).
- [12] Reiter, J. P., Wang, Q., and Zhang, B.: Bayesian estimation of disclosure risks for multiply imputed, synthetic data, *Journal of Privacy and Confidentiality*, Vol. 6, No. 1(2014).
- [13] Stadler, T., Oprisanu, B., and Troncoso, C.: Synthetic data-anonymisation groundhog day, *31st USENIX Security Symposium (USENIX Security 22)*, 2022, pp. 1451–1468.
- [14] Von Elm, E., Altman, D. G., Egger, M., Pocock,



S. J., Göttsche, P. C., Vandembroucke, J. P., Initiative, S., et al.: The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies, *International journal of surgery*, Vol. 12,

No. 12(2014), pp. 1495–1499.

[15] 国立研究開発法人産業技術総合研究所: 機械学習品質マネジメントガイドライン 第3版 (Revision 3.2.1), 技術報告, 国立研究開発法人産業技術総合研究所, 2022.