

(19)日本国特許庁(JP)

(12)特許公報(B1)

(11)特許番号  
特許第7369325号  
(P7369325)

(45)発行日 令和5年10月25日(2023.10.25)

(24)登録日 令和5年10月17日(2023.10.17)

(51)国際特許分類 F I  
G 0 6 N 20/00 (2019.01) G 0 6 N 20/00  
G 0 6 N 3/08 (2023.01) G 0 6 N 3/08

請求項の数 13 (全27頁)

(21)出願番号	特願2023-513648(P2023-513648)	(73)特許権者	399037405 楽天グループ株式会社 東京都世田谷区玉川一丁目14番1号
(86)(22)出願日	令和4年3月24日(2022.3.24)	(74)代理人	110000154 弁理士法人はるか国際特許事務所
(86)国際出願番号	PCT/JP2022/014019	(72)発明者	蔡 永男 東京都世田谷区玉川一丁目14番1号 楽天グループ株式会社内
審査請求日	令和5年2月27日(2023.2.27)	(72)発明者	ブラキャッシャー プリーサム 東京都世田谷区玉川一丁目14番1号 楽天グループ株式会社内
早期審査対象出願		(72)発明者	金 美廷 東京都世田谷区玉川一丁目14番1号 楽天グループ株式会社内
		(72)発明者	ラハ ポウロミ

最終頁に続く

(54)【発明の名称】 学習システム、学習方法、及びプログラム

(57)【特許請求の範囲】

【請求項1】

第1年齢の第1人物に関する第1訓練画像を取得する第1取得部と、  
第2年齢の第2人物に関する第2訓練画像を取得する第2取得部と、  
第3年齢の第3人物に関する第3訓練画像を取得する第3取得部と、  
推定対象画像に含まれる推定対象人物に関する推定対象年齢の推定結果として、所定の  
範囲内の年齢ごとに、当該推定対象人物が当該年齢である確率を示す確率分布を出力する  
学習モデルの学習処理を実行する学習部と、  
を含み、  
前記学習部は、  
前記第1訓練画像を前記学習モデルに入力した場合に前記学習モデルから出力された前  
記確率分布である第1分布に示された前記年齢及び前記確率に基づいて、第1平均年齢を  
計算し、  
前記第2訓練画像を前記学習モデルに入力した場合に前記学習モデルから出力された前  
記確率分布である第2分布に示された前記年齢及び前記確率に基づいて、第2平均年齢を  
計算し、  
前記第3訓練画像を前記学習モデルに入力した場合に前記学習モデルから出力された前  
記確率分布である第3分布に示された前記年齢及び前記確率に基づいて、第3平均年齢を  
計算し、  
前記第1平均年齢、前記第2平均年齢及び前記第3平均年齢のうちのいずれか2つの差

と、他のいずれか2つの差と、の差が、前記第1年齢、前記第2年齢及び前記第3年齢のうちいずれか2つの差と、他のいずれか2つの差と、の差に応じたマージンに近づくように、前記学習処理を実行する、

学習システム。

【請求項2】

前記第1年齢は、前記第2年齢と同じであり、

前記第1年齢は、前記第3年齢とは異なり、

前記学習部は、前記第1平均年齢と前記第2平均年齢との差と、前記第1平均年齢と前記第3平均年齢との差と、の差が、前記第1年齢及び前記第2年齢と、前記第3年齢と、の差に応じた前記マージンに近づくように、前記学習処理を実行する、

請求項1に記載の学習システム。

【請求項3】

前記第1年齢は、前記第2年齢及び前記第3年齢の両方と異なり、

前記学習部は、前記第1平均年齢と前記第2平均年齢との差と、前記第1平均年齢と前記第3平均年齢との差と、の差が、前記第1年齢及び前記第2年齢の差と、前記第1年齢及び前記第3年齢の差と、の差に応じた前記マージンに近づくように、前記学習処理を実行する、

請求項1に記載の学習システム。

【請求項4】

前記第1年齢は、前記第2年齢と同じであり、

前記第1年齢は、前記第3年齢とは異なり、

前記学習部は、

前記第1訓練画像に基づいて、前記学習モデルによる第1処理結果を取得し、

前記第2訓練画像に基づいて、前記学習モデルによる第2処理結果を取得し、

前記第3訓練画像に基づいて、前記学習モデルによる第3処理結果を取得し、

前記第1処理結果と前記第3処理結果との違いと、前記第2処理結果と前記第3処理結果との違いと、に基づいて、前記学習処理を実行する、

請求項1に記載の学習システム。

【請求項5】

前記学習部は、

前記第1訓練画像と、前記学習モデルと、に基づいて、前記第1訓練画像に関する第1特徴量を計算し、

前記第2訓練画像と、前記学習モデルと、に基づいて、前記第2訓練画像に関する第2特徴量を計算し、

前記第1特徴量と、前記第2特徴量と、に基づいて、前記学習処理を実行する、

請求項1～4の何れかに記載の学習システム。

【請求項6】

前記学習部は、

前記第1特徴量と、前記第2特徴量と、に基づいて、コサイン類似度を計算し、

前記コサイン類似度に基づいて、前記学習処理を実行する、

請求項5に記載の学習システム。

【請求項7】

前記学習部は、

前記第1分布と、前記第2分布と、に基づいて、カルバックライブラー情報量を計算し、

前記カルバックライブラー情報量に基づいて、前記学習処理を実行する、

請求項1～6の何れかに記載の学習システム。

【請求項8】

前記学習部は、

前記第1訓練画像に基づいて、前記学習モデルによる第1推定結果を取得し、

前記第1推定結果と、前記第1年齢と、に基づいて、第1損失を計算し、

10

20

30

40

50

前記第 1 訓練画像に基づく前記学習モデルの第 1 処理結果と、前記第 2 訓練画像に基づく前記学習モデルの第 2 処理結果と、の関係に基づいて、関係損失を計算し、  
前記第 1 損失と、前記関係損失と、に基づいて、前記学習処理を実行する、  
請求項 1 ~ 7 の何れかに記載の学習システム。

【請求項 9】

前記学習部は、前記第 1 損失、前記関係損失に関する重み係数、及び前記関係損失に基づいて、前記学習処理を実行する、  
請求項 8 に記載の学習システム。

【請求項 10】

前記学習部は、  
前記第 1 推定結果と、前記第 1 年齢と、に基づいて、複数の前記第 1 損失を計算し、  
前記複数の第 1 損失と、前記関係損失と、に基づいて、前記学習処理を実行する、  
請求項 8 又は 9 に記載の学習システム。

10

【請求項 11】

前記学習部は、  
前記第 1 処理結果と、前記第 2 処理結果と、に基づいて、複数の前記関係損失を計算し、  
前記第 1 損失と、前記複数の関係損失と、に基づいて、前記学習処理を実行する、  
請求項 8 ~ 10 の何れかに記載の学習システム。

【請求項 12】

第 1 年齢の第 1 人物に関する第 1 訓練画像を取得する第 1 取得ステップと、  
第 2 年齢の第 2 人物に関する第 2 訓練画像を取得する第 2 取得ステップと、  
第 3 年齢の第 3 人物に関する第 3 訓練画像を取得する第 3 取得ステップと、  
推定対象画像に含まれる推定対象人物に関する推定対象年齢の推定結果として、所定の  
範囲内の年齢ごとに、当該推定対象人物が当該年齢である確率を示す確率分布を出力する  
学習モデルの学習処理を実行する学習ステップと、  
を含み、

20

前記学習ステップは、  
前記第 1 訓練画像を前記学習モデルに入力した場合に前記学習モデルから出力された前  
記確率分布である第 1 分布に示された前記年齢及び前記確率に基づいて、第 1 平均年齢を  
計算し、  
前記第 2 訓練画像を前記学習モデルに入力した場合に前記学習モデルから出力された前  
記確率分布である第 2 分布に示された前記年齢及び前記確率に基づいて、第 2 平均年齢を  
計算し、  
前記第 3 訓練画像を前記学習モデルに入力した場合に前記学習モデルから出力された前  
記確率分布である第 3 分布に示された前記年齢及び前記確率に基づいて、第 3 平均年齢を  
計算し、

30

前記第 1 平均年齢、前記第 2 平均年齢及び前記第 3 平均年齢のうちのいずれか 2 つの差  
と、他のいずれか 2 つの差と、の差が、前記第 1 年齢、前記第 2 年齢及び前記第 3 年齢の  
うちいずれか 2 つの差と、他のいずれか 2 つの差と、の差に応じたマージンに近づくよう  
に、前記学習処理を実行する、

40

学習方法。

【請求項 13】

第 1 年齢の第 1 人物に関する第 1 訓練画像を取得する第 1 取得部、  
第 2 年齢の第 2 人物に関する第 2 訓練画像を取得する第 2 取得部、  
第 3 年齢の第 3 人物に関する第 3 訓練画像を取得する第 3 取得部、  
推定対象画像に含まれる推定対象人物に関する推定対象年齢の推定結果として、所定の  
範囲内の年齢ごとに、当該推定対象人物が当該年齢である確率を示す確率分布を出力する  
学習モデルの学習処理を実行する学習部、  
としてコンピュータを機能させ、  
前記学習部は、

50

前記第 1 訓練画像を前記学習モデルに入力した場合に前記学習モデルから出力された前記確率分布である第 1 分布に示された前記年齢及び前記確率に基づいて、第 1 平均年齢を計算し、

前記第 2 訓練画像を前記学習モデルに入力した場合に前記学習モデルから出力された前記確率分布である第 2 分布に示された前記年齢及び前記確率に基づいて、第 2 平均年齢を計算し、

前記第 3 訓練画像を前記学習モデルに入力した場合に前記学習モデルから出力された前記確率分布である第 3 分布に示された前記年齢及び前記確率に基づいて、第 3 平均年齢を計算し、

前記第 1 平均年齢、前記第 2 平均年齢及び前記第 3 平均年齢のうちのいずれか 2 つの差と、他のいずれか 2 つの差と、の差が、前記第 1 年齢、前記第 2 年齢及び前記第 3 年齢のうちいずれか 2 つの差と、他のいずれか 2 つの差と、の差に応じたマージンに近づくように、前記学習処理を実行する、

10

プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、学習システム、学習方法、及びプログラムに関する。

【背景技術】

【0002】

従来、画像に含まれるオブジェクトに関する数値を推定する学習モデルが知られている。例えば、非特許文献 1 には、オブジェクトの一例である人間の顔写真から目又は鼻等の部位を示す部分を切り出して、年齢を推定する学習モデルに入力する技術が記載されている。この学習モデルは、年齢ごとの確率を示す確率分布を出力する。この確率分布に示された年齢と確率が乗算された値を合計することによって、年齢が推定される。

20

【0003】

例えば、非特許文献 2 には、顔写真が入力されると年齢ごとのスコアを出力する学習モデルの学習方法として、ソフトマックス関数に関するソフトマックス損失、学習モデルの出力に応じた平均年齢と正解の年齢とのずれに関する平均損失、及び学習モデルの出力に応じた年齢の分散に関する分散損失を利用する技術が記載されている。例えば、非特許文献 3 には、顔写真から複数の領域を特定し、個々の領域を学習モデルに入力することによって、年齢を推定する技術が記載されている。

30

【先行技術文献】

【非特許文献】

【0004】

【文献】Rasmus Rothe, Radu Timofte, Luc Van Gool, “ Deep EXpectation of apparent age from a single image ”, ICCVW2015, [https://openaccess.thecvf.com/content\\_iccv\\_2015\\_workshops/w11/papers/Rothe\\_DEX\\_Deep\\_EXpectation\\_ICCV\\_2015\\_paper.pdf](https://openaccess.thecvf.com/content_iccv_2015_workshops/w11/papers/Rothe_DEX_Deep_EXpectation_ICCV_2015_paper.pdf)

Hongyu Pan, Hu Han, Shiguang Shan, Xilin Chen, “ Mean-Variance Loss for Deep Age Estimation from a Face ”, CVPR2018, [https://openaccess.thecvf.com/content\\_cvpr\\_2018/papers/Pan\\_Mean-Variance\\_Loss\\_for\\_CVPR\\_2018\\_paper.pdf](https://openaccess.thecvf.com/content_cvpr_2018/papers/Pan_Mean-Variance_Loss_for_CVPR_2018_paper.pdf)  
Chao Zhang, Shuaicheng Liu, Xun Xu, Ce Zhu, “ Exploring the Limits of Compact Model for Age Estimation ”, CVPR2019, [https://openaccess.thecvf.com/content\\_CVPR\\_2019/papers/Zhang\\_C3AE\\_Exploring\\_the\\_Limits\\_of\\_Compact\\_Model\\_for\\_Age\\_Estimation\\_CVPR\\_2019\\_paper.pdf](https://openaccess.thecvf.com/content_CVPR_2019/papers/Zhang_C3AE_Exploring_the_Limits_of_Compact_Model_for_Age_Estimation_CVPR_2019_paper.pdf)

40

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、非特許文献 1 - 3 の技術では、学習モデルの精度を高めようとする、

50

多数の訓練画像を準備する必要があるため、非常に手間がかかっていた。この点は、オブジェクトの年齢を推定する学習モデルに限られず、年齢以外の数値（例えば、身長又は体重）を推定する学習モデルについても同様である。このため、少量の訓練画像だったとしても、画像に含まれるオブジェクトに関する数値を推定する学習モデルの精度を高めることが求められる。

【0006】

本開示の目的の1つは、画像に含まれるオブジェクトに関する数値を推定する学習モデルの精度を高めることである。

【課題を解決するための手段】

【0007】

本開示に係る学習システムは、第1数値の第1オブジェクトに関する第1訓練画像を取得する第1取得部と、第2数値の第2オブジェクトに関する第2訓練画像を取得する第2取得部と、前記第1訓練画像及び前記第2訓練画像を利用した距離学習に基づいて、推定対象画像に含まれる推定対象オブジェクトに関する推定対象数値を推定する学習モデルの学習処理を実行する学習部と、を含む。

【発明の効果】

【0008】

本開示によれば、画像に含まれるオブジェクトに関する数値を推定する学習モデルの精度が高まる。

【図面の簡単な説明】

【0009】

【図1】学習システムの全体構成の一例を示す図である。

【図2】学習済みの学習モデルが利用される様子の一例を示す図である。

【図3】学習モデルの学習処理の一例を示す図である。

【図4】学習システムで実現される機能の一例を示す機能ブロック図である。

【図5】訓練データベースの一例を示す図である。

【図6】コサイン類似度損失を利用した手法の一例を示す図である。

【図7】トリプレットマージン損失を利用した手法の一例を示す図である。

【図8】学習システムで実行される処理の一例を示す図である。

【図9】変形例2におけるトリプレットマージン損失の一例を示す図である。

【発明を実施するための形態】

【0010】

[1. 学習システムの全体構成]

本開示に係る学習システムの実施形態の一例を説明する。図1は、学習システムの全体構成の一例を示す図である。例えば、学習システム1は、推定装置10と学習装置20とを含む。本実施形態では、推定装置10と学習装置20とが通信ケーブルCにより通信可能に接続される場合を説明するが、推定装置10と学習装置20とは、インターネット又はLAN等のネットワークにより通信可能に接続されてもよい。学習システム1は、少なくとも1つのコンピュータを含めばよく、図1の例に限られない。例えば、学習システム1は、学習装置20だけを含んでもよい。

【0011】

推定装置10は、学習済みの学習モデルを利用するコンピュータである。例えば、推定装置10は、パーソナルコンピュータ、スマートフォン、タブレット端末、ウェアラブル端末、又はサーバコンピュータである。制御部11は、少なくとも1つのプロセッサを含む。記憶部12は、RAM等の揮発性メモリと、ハードディスク等の不揮発性メモリと、を含む。通信部13は、有線通信用の通信インタフェースと、無線通信用の通信インタフェースと、の少なくとも一方を含む。撮影部14は、少なくとも1つのカメラを含む。

【0012】

学習装置20は、学習済みの学習モデルを作成するコンピュータである。例えば、学習装置20は、パーソナルコンピュータ、スマートフォン、タブレット端末、ウェアラブル

10

20

30

40

50

端末、又はサーバコンピュータである。制御部 2 1、記憶部 2 2、及び通信部 2 3 の物理的構成は、それぞれ制御部 1 1、記憶部 1 2、及び通信部 1 3 と同様であってよい。操作部 2 4 は、キーボード又はマウス等の入力デバイスである。表示部 2 5 は、液晶ディスプレイ又は有機 E L ディスプレイである。

#### 【 0 0 1 3 】

なお、記憶部 1 2、2 2 に記憶されるプログラムは、ネットワークを介して供給されてもよい。また、推定装置 1 0 又は学習装置 2 0 に、コンピュータ読み取り可能な情報記憶媒体を読み取る読取部（例えば、光ディスクドライブやメモリカードスロット）、又は、外部機器とデータの入出力をするための入出力部（例えば、U S B ポート）が含まれてもよい。例えば、情報記憶媒体に記憶されたプログラムが、読取部又は入出力部を介して供給されてもよい。

10

#### 【 0 0 1 4 】

##### [ 2 . 学習システムの概要 ]

図 2 は、学習済みの学習モデルが利用される様子の一例を示す図である。本実施形態では、学習モデルの一例として、ある人間の顔が撮影された画像に基づいて、この人間の年齢を推定する学習モデルを説明する。学習装置 2 0 は、推定装置 1 0 に対し、学習済みの学習モデルを送信する。推定装置 1 0 は、学習装置 2 0 から受信した学習済みの学習モデルを、記憶部 1 2 に記録する。

#### 【 0 0 1 5 】

推定装置 1 0 は、撮影部 1 4 を利用して、年齢の推定対象となる人間である推定対象人間 E H を撮影する。本実施形態では、撮影部 1 4 が動画モードで推定対象人間 E H を撮影する場合を説明するが、撮影部 1 4 は、静止画モードで推定対象人間 E H を撮影してもよい。推定装置 1 0 は、推定対象人間 E H が示された推定対象画像 E I と、学習済みの学習モデルと、に基づいて、推定対象人間 E H の年齢である推定年齢を推定する。推定対象画像 E I には、推定対象人間 E H の顔が示されている。

20

#### 【 0 0 1 6 】

推定年齢は、任意の目的で利用可能である。例えば、推定対象人間 E H が店舗の顧客である場合には、店舗の客層を把握する目的、酒やたばこといった商品の購入時における年齢確認をする目的、又は年齢に応じた広告を提示する目的で、推定年齢が推定されてもよい。他にも例えば、空港又はイベント会場といった施設で本人確認をする目的で、推定年齢が推定されてもよい。

30

#### 【 0 0 1 7 】

学習装置 2 0 は、学習済みの学習モデルを作成するために、訓練画像と、正解となる年齢と、が関連付けられた訓練データを、学習モデルに学習させる。訓練画像は、学習モデルに学習させる画像である。訓練画像には、学習モデルに特徴を学習させる人間が示されている。正解となる年齢は、訓練画像に示された人間の年齢である。本実施形態では、訓練画像と、正解となる年齢と、のペアが訓練データに相当する。

#### 【 0 0 1 8 】

以降、訓練画像に示された人間を、訓練人間という。学習モデルの精度を高めるためには、訓練人間として、多種多様な人間が選出されることが望ましい。しかしながら、この場合、多数の訓練データを用意する必要があるため、非常に手間がかかる。そこで、本実施形態では、距離学習を利用して、訓練データが少なかったとしても、精度の高い学習モデルを作成できるようにしている。

40

#### 【 0 0 1 9 】

図 3 は、学習モデルの学習処理の一例を示す図である。図 3 の例では、3 5 歳の第 1 訓練人間 T H 1 に関する第 1 訓練画像 T I 1、3 5 歳の第 2 訓練人間 T H 2 に関する第 2 訓練画像 T I 2、及び 4 5 歳の第 3 訓練人間 T H 3 に関する第 3 訓練画像 T I 3 が利用される。以降、第 1 訓練人間 T H 1 ~ 第 3 訓練人間 T H 3 を区別しない時は、単に訓練人間 T H という。同様に、第 1 訓練画像 T I 1 ~ 第 3 訓練画像 T I 3 を区別しない時は、単に訓練画像 T I という。訓練画像 T I には、訓練人間 T H の顔が示されている。

50

## 【 0 0 2 0 】

本実施形態では、学習モデルMが畳み込みニューラルネットワークである場合を例に挙げる。図3には、主に、学習モデルMにおける畳み込み層、全結合層、及び出力層の処理が示されている。学習モデルM自体は、種々の機械学習手法を利用可能であり、畳み込みニューラルネットワークに限られない。例えば、再帰型ニューラルネットワーク、ディープニューラルネットワーク、敵対的生成ネットワーク、又は長・短期記憶モデルが利用されてもよい。機械学習手法は、教師有り学習、半教師有り学習、又は教師無し学習の何れの手法も利用可能である。

## 【 0 0 2 1 】

例えば、学習モデルMの畳み込み層は、第1訓練画像T I 1が入力されると、第1訓練画像T I 1に対する畳み込みを実行する。学習モデルMの全結合層は、畳み込み層による計算結果に対する全結合を実行する。学習モデルMの出力層は、活性化関数に基づいて、第1訓練人間T H 1の年齢の推定結果を出力する。学習モデルMは、第2訓練画像T I 2及び第3訓練画像T I 3が入力された場合も同様にして、第2訓練人間T H 2及び第3訓練人間T H 3の年齢の推定結果を出力する。

10

## 【 0 0 2 2 】

本実施形態では、学習モデルMは、年齢の推定結果として、確率分布を出力する。以降、第1訓練画像T I 1が入力された場合に出力される確率分布を、第1分布D 1という。第2訓練画像T I 2が入力された場合に出力される確率分布を、第2分布D 2という。第3訓練画像T I 3が入力された場合に出力される確率分布を、第3分布D 3という。第1分布D 1、第2分布D 2、及び第3分布D 3を区別しない時は、単に確率分布Dという。

20

## 【 0 0 2 3 】

確率分布Dには、ある範囲内の年齢 $j$  ( $j$ は0以上の整数。図3では、0~100の整数。)ごとに、この年齢である確率 $p_j$ が示される。確率 $p_j$ は、蓋然性又はスコアと呼ばれることもある。本実施形態では、全年齢の確率 $p_j$ の合計値が100%であるものとするが、パーセンテージ以外の他の単位で確率 $p_j$ が表現される場合には、全年齢の確率 $p_j$ の合計値は、他の値になってもよい。図3の例では、学習モデルMの精度がある程度高く、どの確率分布Dも、正解となる年齢 $j$ の確率 $p_j$ が最も高いが、学習処理の途中であれば、正解となる年齢 $j$ の確率 $p_j$ が最も高くなるとは限らない。

## 【 0 0 2 4 】

学習装置20は、現状の学習モデルMの処理結果に基づいて、理想的な処理結果が得られるように、学習処理を実行する。本実施形態では、この処理結果として、学習モデルMの中間的な処理結果と、学習モデルMの最終的な処理結果と、の両方が利用される場合を説明する。学習装置20は、現状の処理結果と、理想的な処理結果と、の違いである損失が小さくなるように、学習処理を実行する。

30

## 【 0 0 2 5 】

本実施形態では、損失の一例として、ソフトマックス損失、平均損失、分散損失、コサイン類似度損失(図3では「cos sim loss」と、及びトリプレットマージン損失といった5つの損失を説明する。ソフトマックス損失、平均損失、及び分散損失は、ある1つの訓練画像T Iの中での損失である。コサイン類似度損失及びトリプレットマージン損失は、複数の訓練画像T Iの互いの関係に関する損失である。

40

## 【 0 0 2 6 】

例えば、学習装置20は、第1訓練画像T I 1に基づいて、ソフトマックス損失、平均損失、及び分散損失を計算する。ソフトマックス損失は、第1分布D 1の中で確率 $p_j$ が最も高い年齢 $j$ と、正解となる年齢(図3では、35歳)と、の差に応じた損失である。平均損失は、確率分布Dに基づいて計算される平均年齢と、正解となる年齢と、の差に応じた損失である。分散損失は、年齢 $j$ と平均損失の差に応じた損失である。図3には、各損失の計算式の一例が示されているが、各損失の計算式自体は、種々の計算式を利用可能である。

## 【 0 0 2 7 】

50

例えば、学習装置 20 は、第 1 訓練画像 T I 1 と、第 2 訓練画像 T I 2 と、に基づいて、コサイン類似度損失を計算する。コサイン類似度損失は、学習モデル M の全結合層により計算された第 1 訓練画像 T I 1 の第 1 特徴量 F 1 と、学習モデル M の全結合層により計算された第 2 訓練画像 T I 1 の第 2 特徴量 F 1 と、のコサイン類似度に応じた損失である。本実施形態では、第 1 訓練人間 T H 1 と第 2 訓練人間 T H 2 とが互いに同じ年齢なので、コサイン類似度が高いことが望ましい。このため、コサイン類似度が高いほど、コサイン類似度損失は低くなる。

【 0 0 2 8 】

例えば、学習装置 20 は、第 1 訓練画像 T I 1、第 2 訓練画像 T I 2、及び第 3 訓練画像 T I 3 に基づいて、トリプレットマージン損失を計算する。本実施形態では、第 1 訓練人間 T H 1 と第 2 訓練人間 T H 2 とが互いに同じ年齢なので、第 1 訓練画像 T I 1 に基づく処理結果と、第 2 訓練画像 T I 2 に基づく処理結果と、が似ることが望ましい。逆に、第 1 訓練人間 T H 1 と第 3 訓練人間 T H 3 とが互いに同じ年齢なので、第 1 訓練画像 T I 1 に基づく処理結果と、第 3 訓練画像 T I 3 に基づく処理結果と、が似ないことが望ましい。これら 3 つの処理結果の関係に応じたトリプレットマージン損失が計算される。

10

【 0 0 2 9 】

本実施形態では、ソフトマックス損失、平均損失、分散損失、コサイン類似度損失、及びトリプレットマージン損失が合計された合計損失が小さくなるように、学習処理が実行される。学習システム 1 は、コサイン類似度損失及びトリプレットマージン損失といった距離学習に関する損失を利用することによって、訓練データが少なかったとしても、学習モデル M の精度が高まるようになっている。以降、学習システム 1 の詳細を説明する。

20

【 0 0 3 0 】

[ 3 . 学習システムで実現される機能 ]

図 4 は、学習システム 1 で実現される機能の一例を示す機能ブロック図である。

【 0 0 3 1 】

[ 3 - 1 . 推定装置で実現される機能 ]

データ記憶部 100 は、記憶部 12 を主として実現される。推定部 101 は、制御部 11 を主として実現される。

【 0 0 3 2 】

[ データ記憶部 ]

データ記憶部 100 は、推定対象人間 E H の年齢を推定するために必要なデータを記憶する。例えば、データ記憶部 100 は、学習済みの学習モデル M を記憶する。データ記憶部 100 は、学習済みの学習モデル M 以外にも任意のデータを記憶可能である。例えば、データ記憶部 100 は、推定対象画像 E I と、学習済みの学習モデル M により推定された年齢と、を関連付けて記憶してもよい。

30

【 0 0 3 3 】

[ 推定部 ]

推定部 101 は、推定対象画像 E I と、学習済みの学習モデル M と、に基づいて、推定対象人間 E H の推定年齢を推定する。本実施形態では、推定部 101 が、撮影部 14 により生成された推定対象画像 E I を取得する場合を説明するが、推定部 101 は、推定装置 10 以外の他のコンピュータ又は他の情報記憶媒体から、推定対象画像 E I を取得してもよい。推定部 101 は、学習済みの学習モデル M に、推定対象画像 E I を入力する。

40

【 0 0 3 4 】

例えば、学習モデル M の畳み込み層は、入力された推定対象画像 E I の畳み込みを実行する。学習モデル M の全結合層は、畳み込みの実行結果に対する全結合を実行し、推定対象画像 E I の特徴量である推定対象特徴量を取得する。学習モデル M の出力層は、推定対象特徴量に基づいて、推定年齢の推定結果を出力する。これらの一連の処理は、後述の学習部 204 により調整されたパラメータに基づいて実行される。

【 0 0 3 5 】

本実施形態では、学習モデル M の出力層が、推定結果として確率分布 D を出力し、推定

50

部 1 0 1 が、確率分布 D の中で最も高い確率  $p_j$  の年齢  $j$  を、推定年齢として推定する場合を例に挙げる。推定対象人間 E H の年齢の推定方法は、他の方法であってもよく、本実施形態の例に限られない。例えば、推定部 1 0 1 は、2 番目以降に高い確率  $p_j$  の年齢  $j$  を、推定年齢として推定してもよい。例えば、推定部 1 0 1 は、確率分布 D の年齢  $j$  と確率  $p_j$  に基づいて計算した平均年齢を、推定対象人間 E H の年齢として推定してもよい。

【 0 0 3 6 】

なお、学習モデル M の出力層が出力する推定結果は、確率分布 D 以外の他の推定結果であってもよく、確率分布 D に限られない。推定部 1 0 1 が確率分布 D に基づいて実行するものとして説明した処理が、学習モデル M の出力層によって実行され、当該処理の実行結果が出力されてもよい。例えば、学習モデル M の出力層は、確率分布 D の中で最も高い確率  $p_j$  の年齢  $j$  を特定し、当該特定された年齢  $j$  を出力してもよい。例えば、学習モデル M の出力層は、平均年齢を計算し、当該計算された平均年齢を出力してもよい。学習済みの学習モデル M の出力層を付け替えることによって、このような推定結果を出力可能である。

10

【 0 0 3 7 】

[ 3 - 2 . 学習装置で実現される機能 ]

データ記憶部 2 0 0 は、記憶部 2 2 を主として実現される。第 1 取得部 2 0 1、第 2 取得部 2 0 2、第 3 取得部 2 0 3、及び学習部 2 0 4 は、制御部 2 1 を主として実現される。

【 0 0 3 8 】

[ データ記憶部 ]

データ記憶部 2 0 0 は、学習モデル M の学習処理に必要なデータを記憶する。例えば、データ記憶部 2 0 0 は、学習処理が完了する前の学習モデル M と、訓練データベース D B と、を記憶する。学習処理の開始前であれば、データ記憶部 2 0 0 は、パラメータが初期値の学習モデル M を記憶する。学習処理の実行中であれば、データ記憶部 2 0 0 は、パラメータを調整中の学習モデル M を記憶する。学習処理の完了後であれば、データ記憶部 2 0 0 は、学習済みの学習モデル M を記憶する。

20

【 0 0 3 9 】

データ記憶部 2 0 0 は、互いにパラメータを共有する複数の学習モデル M を記憶してもよい。例えば、図 3 の処理は、1 つの学習モデル M に基づいて実行されるのではなく、複数の学習モデル M に基づいて実行されてもよい。第 1 学習モデル M 1、第 2 学習モデル M 2、及び第 3 学習モデル M 3 が互いにパラメータを共有したとすると、第 1 訓練画像 T I 1 が第 1 学習モデル M 1 に入力され、第 2 訓練画像 T I 2 が第 2 学習モデル M 2 に入力され、第 3 訓練画像 T I 3 が第 3 学習モデル M 3 に入力されてもよい。第 1 学習モデル M 1、第 2 学習モデル M 2、及び第 3 学習モデル M 3 の何れかのパラメータが学習処理によって変わった場合には、残りの 2 つのパラメータも同じ値になる。

30

【 0 0 4 0 】

本実施形態では、学習モデル M が畳み込みニューラルネットワークなので、データ記憶部 2 0 0 は、畳み込み層、全結合層、及び出力層を含む学習モデル M のプログラムを記憶する。学習モデル M のパラメータは、学習モデル M の各層（プログラム部分）とデータとして一体であってもよいし別々であってもよい。畳み込みニューラルネットワーク以外の他の機械学習手法を利用する場合も、データ記憶部 2 0 0 は、他の機械学習手法に応じた形式の学習モデル M を記憶すればよい。

40

【 0 0 4 1 】

図 5 は、訓練データベース D B の一例を示す図である。訓練データベース D B は、複数の訓練データが格納されたデータベースである。例えば、訓練データは、ある訓練人間 T H に関する訓練画像 T I と、この訓練人間 T H の年齢と、のペアを含む。個々の訓練画像 T I の形式は、互いに同じものとするが、多少であれば異なってもよい。ここでの形式とは、画像データとしての形式である。例えば、形式は、拡張子、解像度（ピクセル数）、縦横比、色数、又はこれらの組み合わせである。訓練画像 T I には、訓練人間 T H の顔の一部だけが示されてもよい。

50

## 【 0 0 4 2 】

本実施形態では、学習モデルMの作成者が訓練データベースDBを作成する場合を説明するが、訓練データの作成方法自体は、種々の方法を利用可能である。クラスタリング等を利用して訓練データの作成を自動化する方法も存在するので、このような方法を利用して訓練データの作成が自動化されてもよい。訓練画像TIには、後述する第1オブジェクト等に応じたオブジェクトが示されていればよい。このため、訓練画像TIには、人間以外の他のオブジェクトが示されていてもよい。

## 【 0 0 4 3 】

本実施形態では、訓練データベースDBに格納された訓練画像TIの用途が指定されないものとする。ここでの用途とは、訓練画像TIが、第1訓練画像TI1、第2訓練画像TI2、又は第3訓練画像TI3の何れとして利用されるかである。本実施形態のように、訓練画像TIの用途が限定されない場合、訓練画像TIは、第1訓練画像TI1、第2訓練画像TI2、及び第3訓練画像TI3の何れにもなりうる。

10

## 【 0 0 4 4 】

なお、学習モデルMの作成者は、個々の訓練画像TIの用途を指定してもよい。この場合、訓練データベースDBには、作成者が指定した用途を識別可能な情報が格納される。用途ごとに、別々のデータベースが用意されていてもよい。例えば、第1訓練画像TI1だけが格納された第1データベース、第2訓練画像TI2だけが格納された第2データベース、及び第3訓練画像TI3だけが格納された第3データベースが用意されてもよい。他にも例えば、訓練人間THの年齢ごとに、別々のデータベースが用意されてもよい。

20

## 【 0 0 4 5 】

## [ 第1取得部 ]

第1取得部201は、第1年齢の第1訓練人間TH1に関する第1訓練画像TI1を取得する。第1取得部201は、訓練データベースDBの中から、任意の第1訓練画像TI1を取得可能である。例えば、第1取得部201は、ランダムに第1訓練画像TI1を取得してもよい。第1取得部201は、まだ学習モデルMに学習させていない年齢の第1訓練画像TI1を取得してもよい。第1取得部201は、学習モデルMに学習させた枚数が相対的に少ない年齢の第1訓練画像TI1を取得してもよい。

## 【 0 0 4 6 】

第1年齢は、第1訓練人間TH1の年齢である。第1年齢は、第1数値の一例である。このため、第1年齢と記載した箇所は、第1数値と読み替えることができる。第1数値は、第1訓練人間TH1に関する数値である。第1数値は、年齢に限られず、第1人間Hに関する何らかの数値であればよい。第1数値は、第1人間Hの特徴を表す数値であってよい。例えば、第1数値は、第1人間Hの身長、体重、部位（例えば、頭、足、胴）のサイズ、又は体型であってもよい。本実施形態では、第1数値が後述の第2数値と同じである場合を例に挙げる。更に、第1数値が後述の第3数値とは異なる場合を例に挙げる。

30

## 【 0 0 4 7 】

第1訓練人間TH1は、第1訓練画像TI1に示された人間である。第1訓練人間TH1は、第1オブジェクトの一例である。このため、第1訓練人間TH1と記載した箇所は、第1オブジェクトと読み替えることができる。第1オブジェクトは、第1訓練画像TI1に示されたオブジェクトである。第1訓練画像TI1がカメラにより生成された撮影画像であれば、第1オブジェクトは、現実空間に配置された被写体である。第1訓練画像TI1がCGであれば、第1オブジェクトは、仮想空間に配置された3Dオブジェクト、又は、2次元的に描かれたキャラクター等のオブジェクトである。第1訓練画像TI1には、第1訓練人間THの顔の一部だけが示されてもよい。この点は、第2訓練画像TI2及び第3訓練画像TI3も同様である。

40

## 【 0 0 4 8 】

なお、第1オブジェクトは、人間以外の他のオブジェクトであってもよい。例えば、第1オブジェクトは、犬又は猫といった他の動物であってもよい。この場合、第1訓練画像TI1には、他の動物が示される。第1訓練画像TI1には、他の動物の年齢が関連付け

50

られる。学習モデルMは、他の動物の年齢を推定する。第1オブジェクトが他の動物の場合にも、学習モデルMは、年齢ではなく、他の動物の体長、体重、サイズ、又は体型といった他の数値を推定してもよい。

【0049】

また、第1オブジェクトは、動物以外の他のオブジェクトであってもよい。例えば、第1オブジェクトは、植物、家具、室内の壁、飲食物、乗物、建造物、又は自然界のその他の風景であってもよい。この場合、第1訓練画像T I 1には、これらの他のオブジェクトが含まれる。植物又は建造物のように、年齢に相当する概念（樹齢又は築年数）が存在する他のオブジェクトであれば、学習モデルMは、この概念を推定する。このような概念が存在しない他のオブジェクトであれば、学習モデルMは、他のオブジェクトの重量又はサイズ等といった他の数値を推定してもよい。

10

【0050】

[第2取得部]

第2取得部202は、第2年齢の第2訓練人間T H 2に関する第2訓練画像T I 2を取得する。第2取得部202は、訓練データベースD Bの中から、任意の第2訓練画像T I 2を取得可能である。本実施形態では、第2取得部202は、訓練データベースD Bの中から、第1訓練画像T I 1と一緒に距離学習で利用するのに適した第2訓練画像T I 2を取得する場合を説明するが、第2取得部202は、第1訓練画像T I 1に関係なく、ランダムに第2訓練画像T I 2を取得してもよい。

【0051】

第2年齢は、第2訓練人間T H 2の年齢である。第2年齢は、第2数値の一例である。このため、第2年齢と記載した箇所は、第2数値と読み替えることができる。第2数値が年齢に限られない点は、第1数値と同様である。

20

【0052】

第2訓練人間T H 2は、第2訓練画像T I 2に示された人間である。第2訓練人間T H 2は、第2オブジェクトの一例である。このため、第2訓練人間T H 2と記載した箇所は、第2オブジェクトと読み替えることができる。第2オブジェクトは、第2訓練画像T I 2に示されたオブジェクトである。第2オブジェクトが人間に限られない点は、第1オブジェクトと同様である。

【0053】

本実施形態では、第1オブジェクト及び第2オブジェクトは、互いに異なる訓練人間T Hである場合を説明するが、第1オブジェクト及び第2オブジェクトは、互いに同じ訓練人間T Hであってもよい。例えば、第1訓練画像T I 1には、ある訓練人間T Hがある表情で撮影されており、第2訓練画像T I 2には、この訓練人間T Hが別の表情で撮影されていてもよい。他にも例えば、第1訓練画像T I 1には、ある訓練人間T Hがある角度から撮影されており、第2訓練画像T I 2には、この訓練人間T Hが別の角度から撮影されていてもよい。

30

【0054】

本実施形態では、第2取得部202は、ある第1年齢の第1訓練人間T H 1が示された第1訓練画像T I 1が取得された場合に、訓練データベースD Bの中から、この第1年齢と同じ第2年齢の訓練画像T Iを検索し、当該訓練画像T Iを、第2訓練画像T I 2として取得する。第2取得部202は、検索で複数の訓練画像T Iがヒットした場合には、複数の訓練画像T Iのうちの何れかを、第2訓練画像T I 2として取得すればよい。

40

【0055】

なお、第1訓練画像T I 1が取得される前に、作成者により第2年齢が指定されるのであれば、第2取得部202は、第1訓練画像T I 1が取得される前に、第2訓練画像T I 2を取得してもよい。他にも例えば、訓練データベースD Bに、距離学習と一緒に利用する第1訓練画像T I 1及び第2訓練画像T I 2の組み合わせが予め関連付けられているのであれば、第2取得部202は、第1訓練画像T I 1に関連付けられた第2訓練画像T I 2を取得してもよい。

50

## 【 0 0 5 6 】

## [ 第 3 取得部 ]

第 3 取得部 2 0 3 は、第 3 年齢の第 3 訓練人間 T H 3 に関する第 3 訓練画像 T I 3 を取得する。第 3 取得部 2 0 3 は、訓練データベース D B の中から、任意の第 3 訓練画像 T I 3 を取得可能である。本実施形態では、第 3 取得部 2 0 3 は、訓練データベース D B の中から、第 1 訓練画像 T I 1 及び第 2 訓練画像 T I 2 と一緒に距離学習で利用するのに適した第 3 訓練画像 T I 3 を取得する場合を説明するが、第 3 取得部 2 0 3 は、第 1 訓練画像 T I 1 及び第 2 訓練画像 T I 2 に関係なく、ランダムに第 3 訓練画像 T I 3 を取得してもよい。

## 【 0 0 5 7 】

第 3 年齢は、第 3 訓練人間 T H 3 の年齢である。第 3 年齢は、第 3 数値の一例である。このため、第 3 年齢と記載した箇所は、第 3 数値と読み替えることができる。第 3 数値が年齢に限られない点は、第 1 数値及び第 2 数値と同様である。

## 【 0 0 5 8 】

第 3 訓練人間 T H 3 は、第 3 訓練画像 T I 3 に示された人間である。第 3 訓練人間 T H 3 は、第 3 オブジェクトの一例である。このため、第 3 訓練人間 T H 3 と記載した箇所は、第 3 オブジェクトと読み替えることができる。第 3 オブジェクトは、第 3 訓練画像 T I 3 に示されたオブジェクトである。第 3 オブジェクトが人間に限られない点は、第 1 オブジェクト及び第 2 オブジェクトと同様である。

## 【 0 0 5 9 】

本実施形態では、第 3 オブジェクトと、第 1 オブジェクト及び第 2 オブジェクトと、が互いに異なる訓練人間 T H である場合を説明するが、第 3 オブジェクトと、第 1 オブジェクト及び第 2 オブジェクトと、が互いに同じ訓練人間 T H であってもよい。例えば、第 1 訓練画像 T I 1 又は第 2 訓練画像 T I 2 には、ある訓練人間 T H の昔の姿（例えば、2 0 歳の時の姿）が撮影されており、第 3 訓練画像 T I 2 には、この訓練人間 T H の最近の姿（例えば、4 0 歳の時の姿）が撮影されていてもよい。

## 【 0 0 6 0 】

例えば、第 3 取得部 2 0 3 は、ある第 1 年齢の第 1 訓練人間 T H 1 が示された第 1 訓練画像 T I 1 が取得された場合に、訓練データベース D B の中から、この第 1 年齢とは異なる第 3 年齢の訓練画像 T I を検索し、当該訓練画像 T I を、第 3 訓練画像 T I 3 として取得する。第 3 取得部 2 0 3 は、検索で複数の訓練画像 T I がヒットした場合には、複数の訓練画像 T I のうちの何れかを、第 3 訓練画像 T I 2 として取得すればよい。

## 【 0 0 6 1 】

本実施形態では、第 1 年齢と第 3 年齢との差が 1 0 歳で固定されているものとするが、第 1 年齢と第 3 年齢との差は、動的に変化してもよい。例えば、第 1 年齢が若いほど、顔の特徴が加齢に応じて大きく変わるのであれば、第 1 年齢と第 3 年齢との差が小さくなくてもよい。逆に、第 1 年齢が若いほど、顔の特徴が加齢に応じてあまり変わらないのであれば、第 1 年齢と第 3 年齢との差が大きくなってもよい。第 3 取得部 2 0 3 は、第 1 訓練画像 T I 1 の第 1 年齢に基づいて第 3 年齢を決定し、当該決定された第 3 年齢の第 3 訓練画像 T I 3 を検索すればよい。

## 【 0 0 6 2 】

なお、第 1 訓練画像 T I 1 が取得される前に、作成者により第 3 年齢が指定されるのであれば、第 3 取得部 2 0 3 は、第 1 訓練画像 T I 1 が取得される前に、第 3 訓練画像 T I 3 を取得してもよい。他にも例えば、訓練データベース D B に、距離学習と一緒に利用する第 1 訓練画像 T I 1、第 2 訓練画像 T I 2、及び第 3 訓練画像 T I 3 の組み合わせが予め関連付けられているのであれば、第 3 取得部 2 0 3 は、第 1 訓練画像 T I 1 及び第 2 訓練画像 T I 2 に関連付けられた第 3 訓練画像 T I 3 を取得してもよい。

## 【 0 0 6 3 】

## [ 学習部 ]

学習部 2 0 4 は、第 1 訓練画像 T I 1 及び第 2 訓練画像 T I 2 を利用した距離学習に基

10

20

30

40

50

づいて、推定対象画像 E I に含まれる推定対象人間 E H に関する推定年齢を推定する学習モデル M の学習処理を実行する。

【 0 0 6 4 】

距離学習は、複数の訓練データの互いの関係に基づく学習手法である。例えば、距離学習では、同じ又は似た訓練データ同士が近づくように学習処理が実行される。例えば、距離学習では、似ていない訓練データ同士が離れるように学習処理が実行される。本実施形態では、距離学習が、これらの両方の手法を含む意味である場合を説明するが、距離学習は、何れか一方の手法だけを含む意味だけでもよい。

【 0 0 6 5 】

距離学習自体は、種々の手法を利用可能であり、例えば、ユークリッド距離系の手法、マハラノビス距離系の手法、又は角度系の手法を利用可能である。学習部 2 0 4 は、深層距離学習の手法を利用して、学習モデル M の学習処理を実行してもよい。先述したように、本実施形態の学習部 2 0 4 は、学習モデル M の学習処理を実行する。

10

【 0 0 6 6 】

推定対象人間 E H は、推定対象オブジェクトの一例である。このため、推定対象人間 E H と記載した箇所は、推定対象オブジェクトと読み替えることができる。推定対象オブジェクトは、推定対象画像 E I に示されたオブジェクトである。推定対象オブジェクトが人間に限られない点は、第 1 オブジェクト、第 2 オブジェクト、及び第 3 オブジェクトと同様である。

【 0 0 6 7 】

推定年齢は、推定対象数値の一例である。このため、推定年齢と記載した箇所は、推定対象数値と読み替えることができる。推定対象数値が年齢に限られない点は、第 1 数値 ~ 第 3 数値と同様である。

20

【 0 0 6 8 】

学習処理は、訓練データに基づいて、学習モデル M のパラメータを調整する処理である。本実施形態では、学習モデル M が畳み込みニューラルネットワークなので、学習処理は、重み係数又はバイアスといったパラメータを調整する処理である。学習処理自体は、学習モデル M として利用する機械学習手法に応じた処理であればよく、本実施形態の例に限られない。学習処理では、機械学習手法に応じたパラメータが調整されるようにすればよい。

30

【 0 0 6 9 】

本実施形態では、学習部 2 0 4 は、第 1 訓練画像 T I 1、第 2 訓練画像 T I 2、及び第 3 訓練画像 T I 3 を利用した距離学習に基づいて、学習処理を実行する場合を説明するが、学習部 2 0 4 は、第 3 訓練画像 T I 3 を利用せずに、第 1 訓練画像 T I 1 及び第 2 訓練画像 T I 2 に基づいて、学習処理を実行してもよい。学習部 2 0 4 は、2 枚以上の訓練画像 T I に基づいて、学習処理を実行すればよい。例えば、学習部 2 0 4 は、4 枚以上の訓練画像 T I に基づいて、学習処理を実行してもよい。

【 0 0 7 0 】

学習部 2 0 4 は、距離学習以外の他の手法を併用して、学習処理を実行してもよい。本実施形態では、他の手法の一例として、ソフトマックス損失、平均損失、及び分散損失を利用した手法を説明する。他の手法自体は、種々の手法を利用可能であり、本実施形態の例に限られない。距離学習の一例として、コサイン類似度損失を利用した手法と、トリプレットマージン損失を利用した手法と、を説明する。

40

【 0 0 7 1 】

[ ソフトマックス損失、平均損失、及び分散損失を利用した手法 ]

まず、ソフトマックス損失、平均損失、及び分散損失を利用した手法を説明する。これら 3 つの損失は、第 1 損失の一例である。このため、ソフトマックス損失、平均損失、又は分散損失と記載した箇所は、第 1 損失と読み替えることができる。第 1 損失は、ある訓練画像 T I に基づく現状の学習モデル M の処理結果と、理想的な処理結果と、の違いに関する損失である。ある訓練画像 T I で第 1 損失を計算するにあたり、他の訓練画像 T I は

50

利用されない。

【 0 0 7 2 】

例えば、学習部 2 0 4 は、第 1 訓練画像 T I 1 に基づいて、学習モデル M による第 1 処理結果を取得する。第 1 処理結果は、第 1 訓練画像 T I 1 が学習モデル M に入力された場合に実行される処理の結果である。ここでは、第 1 処理結果が学習モデル M による第 1 推定結果である場合を説明するが、第 1 処理結果は、学習モデル M の内部的な計算結果であってもよい。即ち、第 1 処理結果が出力層からの出力である場合を説明するが、第 1 処理結果は、中間層からの出力であってもよい。

【 0 0 7 3 】

第 1 推定結果は、第 1 訓練画像 T I 1 が学習モデル M に入力された場合に、学習モデル M から出力された推定年齢である。本実施形態では、第 1 推定結果は、複数の年齢  $j$  の各々と、第 1 訓練人間 T H 1 が当該年齢である第 1 確率  $p_j$  と、を含む第 1 分布 D 1 である。複数の年齢  $j$  は、複数の数値の一例である。このため、複数の年齢  $j$  について説明している箇所は、複数の数値と読み替えることができる。なお、最も確率  $p_j$  が高い年齢  $j$ 、第 1 分布 D 1 に応じた平均年齢、又は第 1 分布 D 1 に応じた平均分散は、第 1 推定結果に相当してもよい。第 1 推定結果は、これらの例に限られず、例えば、確率  $p_j$  が 2 番目以降の年齢  $j$  が第 1 推定結果に相当してもよい。

10

【 0 0 7 4 】

本実施形態では、学習部 2 0 4 は、第 1 推定結果である第 1 分布 D 1 と、第 1 年齢と、に基づいて、複数の第 1 損失を計算する。例えば、学習部 2 0 4 は、第 1 推定結果である第 1 分布 D 1 に基づいて、最も確率  $p_j$  が高い年齢  $j$  を特定する。学習部 2 0 4 は、当該特定された年齢  $j$  と、第 1 年齢と、の差を、ソフトマックス損失として取得する。学習部 2 0 4 は、第 1 分布 D 1 と、図 3 に示す計算式と、に基づいて、平均年齢を計算する。学習部 2 0 4 は、当該計算された平均年齢と、第 1 年齢と、の差を、平均損失として取得する。学習部 2 0 4 は、第 1 分布 D 1 と、図 3 に示す計算式と、に基づいて、平均分散を計算する。学習部 2 0 4 は、当該計算された平均分散と、第 1 年齢に応じた理想的な分散と、の差を、分散損失として取得する。

20

【 0 0 7 5 】

[ コサイン類似度損失を利用した手法 ]

次に、コサイン類似度損失を利用した手法を説明する。学習部 2 0 4 は、第 2 訓練画像 T I 2 に基づいて、学習モデル M による第 2 処理結果を取得する。第 2 処理結果は、第 2 訓練画像 T I 2 が学習モデル M に入力された場合に実行される処理の結果である。第 2 処理結果は、第 1 処理結果と同様、学習モデル M の内部的な計算結果であってもよいし、学習モデル M から出力された第 2 推定結果であってもよい。コサイン類似度損失の計算では、第 1 処理結果及び第 2 処理結果が、ともに学習モデル M の内部的な計算結果である場合を例に挙げる。

30

【 0 0 7 6 】

なお、第 2 推定結果は、第 2 訓練画像 T I 2 が学習モデル M に入力された場合に、学習モデル M が推定した年齢である。本実施形態では、第 2 推定結果は、複数の年齢  $j$  の各々と、第 2 訓練人間 T H 2 が当該年齢である第 2 確率  $p_j$  と、を含む第 2 分布 D 2 である。第 2 推定結果は、第 1 推定結果と同様に、確率分布 D 以外の他の推定結果であってもよい。他の推定結果としては、第 1 推定結果の説明で例示した内容と同様であってもよい。

40

【 0 0 7 7 】

図 6 は、コサイン類似度損失を利用した手法の一例を示す図である。例えば、学習部 2 0 4 は、第 1 訓練画像 T I 1 と、学習モデル M と、に基づいて、第 1 訓練画像 T I 1 に関する第 1 特徴量 F 1 を計算する。例えば、学習部 2 0 4 は、第 2 訓練画像 T I 2 と、学習モデル M と、に基づいて、第 2 訓練画像 T I 2 に関する第 2 特徴量 F 2 を計算する。図 6 では、原点 O の多次元ベクトル空間上で、第 1 特徴量 F 1 及び第 2 特徴量 F 2 が示されている。

【 0 0 7 8 】

50

第1特徴量F1は、第1訓練画像TI1の特徴に関する情報である。第2特徴量F2は、第2訓練画像TI2の特徴に関する情報である。本実施形態では、第1特徴量F1及び第2特徴量F2が多次元ベクトルで表現される場合を説明するが、第1特徴量F1及び第2特徴量F2は、任意の形式で表現可能である。例えば、第1特徴量F1及び第2特徴量F2は、配列又は単一の数値といった他の形式で表現されてもよい。

【0079】

本実施形態では、学習モデルMが畳み込みニューラルネットワークである場合を説明するので、学習部204は、現状の学習モデルMのパラメータに基づいて、第1訓練画像TI1を畳み込むことによって、第1特徴量F1を計算する。学習部204は、現状の学習モデルMのパラメータに基づいて、第2訓練画像TI2を畳み込むことによって、第2特徴量F2を計算する。第1特徴量F1及び第2特徴量F2の計算方法は、図3を参照して説明した通りである。第1特徴量F1及び第2特徴量F2の計算方法自体は、学習モデルMとして利用する機械学習手法に応じた計算方法であればよく、種々の方法を利用可能である。

10

【0080】

例えば、学習部204は、第1特徴量F1と、第2特徴量F2と、に基づいて、学習処理を実行する。本実施形態では、学習部204は、第1特徴量F1と、第2特徴量F2と、に基づいて、コサイン類似度を計算する。学習部204は、このコサイン類似度に基づいて、学習処理を実行する。学習部204は、コサイン類似度に応じたコサイン類似度損失が小さくなるように（コサイン類似度が大きくなるように）、学習処理を実行する。コサイン類似度損失の計算式は、図6の通りである。即ち、学習部204は、第1処理結果と第2処理結果の違いが小さくなるように、学習処理を実行する。

20

【0081】

コサイン類似度損失は、関係損失の一例である。このため、コサイン類似度損失について説明している箇所は、関係損失と読み替えることができる。関係損失は、複数の訓練画像TIの関係に関する損失である。学習部204は、第1訓練画像TI1に基づく学習モデルMの第1処理結果と、第2訓練画像TI2に基づく学習モデルMの第2処理結果と、の関係に基づいて、関係損失を計算する。関係損失自体は、任意の損失であってよく、コサイン類似度損失に限られない。例えば、関係損失は、ユークリッド距離系の手法で計算された損失、又は、マハラノビス距離系の手法で計算された損失であってもよい。

30

【0082】

なお、学習部204は、第3訓練画像TI3と、学習モデルMと、に基づいて、第3訓練画像TI3に関する第3特徴量を計算してもよい。この場合、学習部204は、第1特徴量と、第3特徴量と、に基づいて、コサイン類似度を計算し、このコサイン類似度に基づいて、学習処理を実行してもよい。即ち、学習部204は、第1処理結果と第3処理結果の違いが大きくなるように、学習処理を実行してもよい。上記を組み合わせると、学習部204は、第1処理結果と第2処理結果との違いが小さくなり、かつ、第1処理結果と第3処理結果との違いが大きくなるように、学習処理を実行してもよい。

【0083】

学習部204は、第2特徴量と、第3特徴量と、に基づいて、コサイン類似度を計算し、このコサイン類似度に基づいて、学習処理を実行してもよい。学習部204は、このコサイン類似度が低くなるように、学習処理を実行してもよい。即ち、学習部204は、第2処理結果と第3処理結果の違いが大きくなるように、学習処理を実行してもよい。上記を組み合わせると、学習部204は、第1処理結果と第2処理結果との違いが小さくなり、かつ、第1処理結果と第3処理結果との違いが大きくなるように、かつ、第2処理結果と第3処理結果との違いが大きくなるように、学習処理を実行してもよい。

40

【0084】

[トリプレットマージン損失]

最後に、トリプレットマージン損失を利用した手法を説明する。学習部204は、第3訓練画像TI3に基づいて、学習モデルMによる第3処理結果を取得する。第3処理結果

50

は、第3訓練画像T I 3が学習モデルMに入力された場合に実行される処理の結果である。ここでは、第3処理結果が学習モデルMによる第3推定結果である場合を説明するが、第3処理結果は、学習モデルMの内部的な計算結果であってもよい。即ち、第3処理結果が出力層からの出力である場合を説明するが、第3処理結果は、中間層からの出力であってもよい。

【0085】

第3推定結果は、第3訓練画像T I 3が学習モデルMに入力された場合に、学習モデルMが推定した年齢である。本実施形態では、第3推定結果は、複数の年齢jの各々と、第3訓練人間T H 3が当該年齢jである第3確率 $p_j$ と、を含む第3分布D 3である。第3推定結果は、第1推定結果及び第2推定結果と同様に、確率分布D以外の他の推定結果であってよい。他の推定結果としては、第1推定結果の説明で例示した内容と同様であってよい。本実施形態では、トリプレットマージン損失は、第1推定結果、第2推定結果、及び第3推定結果に基づいて計算されるものとする。

10

【0086】

図7は、トリプレットマージン損失を利用した手法の一例を示す図である。本実施形態では、第1訓練画像T I 1は、トリプレットマージン損失を計算する際の基準となるアンカー画像に相当する。第2訓練画像T I 2は、第1年齢と同じ第2年齢が関連付けられているので、アンカー画像と同じラベルが付与されたポジティブ画像に相当する。第3訓練画像T I 3は、第1年齢とは異なる第3年齢が関連付けられているので、アンカー画像とは異なるラベルが付与されたネガティブ画像に相当する。トリプレットマージン損失を利用した手法では、アンカー画像とポジティブ画像が近づき、かつ、アンカー画像とネガティブ画像が遠ざかるように、学習処理が実行される。

20

【0087】

図7では、第1分布D 1から計算された第1平均年齢 $A A 1$ 、第2分布D 2から計算された第2平均年齢 $A A 2$ 、及び第3分布D 3から計算された第3平均年齢 $A A 3$ が示されている。第2平均年齢 $A A 2$ 及び第3辺金年齢 $A A 3$ の計算方法自体は、第1平均年齢 $A A 1$ と同様である。学習部204は、第1平均年齢 $A A 1$ と第2平均年齢 $A A 2$ との違い $d p$ が小さくなり、かつ、第1平均年齢 $A A 1$ と第3平均年齢 $A A 3$ との違い $d n$ が大きくなるように、学習処理を実行する。

【0088】

例えば、学習部204は、違い $d p$ 、違い $d n$ 、及び所定の計算式に基づいて、トリプレットマージン損失を計算する。この計算式の一例は、図7に示す通りである。計算式自体は、トリプレットマージン損失で利用される他の計算式であってもよい。例えば、トリプレットマージン損失におけるマージンを示すハイパーパラメータを  $\text{margin}$  とする。理想的には、違い $d n$ と違い $d p$ の差がマージン になるので、現状の違い $d n$ と違い $d p$ の差と、マージン と、の差がトリプレットマージン損失になる。

30

【0089】

例えば、学習部204は、トリプレットマージン損失が小さくなるように学習処理を実行することによって、違い $d n$ と違い $d p$ との差がマージン に近づくように、学習処理を実行することになる。図7の例では、第1年齢及び第2年齢と、第3年齢と、の差が10歳なので、マージン としては、10程度の値になる。マージン は、固定値であってもよいし、動的に決定されてもよい。

40

【0090】

学習部204は、ソフトマックス損失、平均損失、分散損失、コサイン類似度損失、及びトリプレットマージン損失を合計し、合計損失を計算する。学習部204は、合計損失が小さくなるように、学習処理を実行する。損失に応じた学習処理自体は、種々の手法を利用可能である。例えば、誤差逆伝播法又は勾配降下法といった手法が利用されてもよい。本実施形態では、合計損失が単純な合計値である場合を説明するが、後述の変形例のように、重み付け係数が加味されてもよい。

【0091】

50

本実施形態では、ソフトマックス損失、平均損失、及び分散損失は第1損失に相当し、コサイン類似度損失及びトリプレットマージン損失は関係損失に相当するので、学習部204は、第1損失と、関係損失と、に基づいて、学習処理を実行することになる。更に、ソフトマックス損失、平均損失、及び分散損失といった複数の第1損失が利用されるので、学習部204は、複数の第1損失と、関係損失と、に基づいて、学習処理を実行することになる。

【0092】

以上のように、本実施形態では、トリプレットマージン損失を利用することによって、学習部204は、第1推定結果、第2推定結果、及び第3推定結果に基づいて、学習処理を実行する。例えば、学習部204は、第1推定結果と第2推定結果との違いが小さくなり、かつ、第1推定結果と第3推定結果との違いが大きくなるように、学習処理を実行する。コサイン類似度損失以外の各損失は、確率分布Dが利用されるので、学習部204は、第1分布D1、第2分布D2、及び第3分布D3に基づいて、学習処理を実行することになる。

10

【0093】

なお、学習部204は、学習モデルMの推定結果ではなく、中間的な処理結果を利用して、トリプレットマージン損失を計算してもよい。即ち、学習部204は、第1処理結果、第2処理結果、及び第3処理結果に基づいて、学習処理を実行してもよい。例えば、学習部204は、第1処理結果と第2処理結果との違いが小さくなり、かつ、第1処理結果と第3処理結果との違いが大きくなるように、学習処理を実行してもよい。

20

【0094】

例えば、学習部204は、第3訓練画像TI3と、学習モデルMと、に基づいて、第3訓練画像TI3に関する第3特徴量F3を計算する。第3特徴量F3は、第3訓練画像TI3の特徴に関する情報である。第3特徴量F3は、第1特徴量F1及び第2特徴量F2と同様、任意の形式で表現可能である。学習部204は、第1特徴量F1と第3特徴量F3との違い $d_n$ と、第1特徴量F1と第2特徴量F2との違い $d_p$ と、の差がマージンに近づくように、学習処理を実行してもよい。

【0095】

例えば、学習部204は、第1処理結果と第3処理結果との違いが、第1年齢と第3年齢との差に応じた違いになるように、学習処理を実行してもよい。学習部204は、第1年齢と第3年齢の差に基づいて、マージンを決定してもよい。学習部204は、第1年齢と第3年齢の差が大きいほど、マージンが大きくなるように、マージンを決定する。学習部204は、当該決定されたマージンに基づいて、学習処理を実行する。図7の例では、第1年齢と第3年齢の差が10歳であったが、第1年齢と第3年齢の差が5歳であれば、マージンは5歳程度の差であってもよい。

30

【0096】

なお、学習部204は、第1損失の一例であるソフトマックス損失、平均損失、及び分散損失を利用せずに、関係損失の一例であるコサイン類似度損失及びトリプレットマージン損失に基づいて、学習処理を実行してもよい。学習部204は、コサイン類似度損失又はトリプレットマージン損失の何れか一方に基づいて、学習処理を実行してもよい。学習部204は、ある1つの関係損失だけに基づいて、学習処理を実行してもよい。

40

【0097】

[4. 学習システム1で実行される処理]

図8は、学習システム1で実行される処理の一例を示す図である。図8では、学習装置20の処理が示されている。学習装置20は、訓練データベースDBに基づいて、第1訓練画像TI1を取得する(S1)。S1では、学習装置20は、訓練データベースDBの中から、学習モデルMにまだ学習させていない訓練画像TIを、第1訓練画像TI1として取得する。

【0098】

学習装置20は、現状の学習モデルMと、第1訓練画像TI1と、に基づいて、第1特

50

微量F 1と、第1分布D 1と、を取得する(S 2)。S 2では、学習装置2 0は、学習モデルMに第1訓練画像T I 1を入力し、学習モデルMの各層に応じた計算を実行する。学習装置2 0は、全結合層により計算された第1特徴量F 1と、出力層により出力された第1分布D 1と、を取得する。

**【0 0 9 9】**

学習装置2 0は、第1訓練画像T I 1に関連付けられた第1年齢と、S 2で取得された第1分布D 1と、に基づいて、第1平均年齢A A 1、ソフトマックス損失、平均損失、及び分散損失を取得する(S 3)。これらの計算方法は、先述した通りである。学習装置2 0は、訓練データベースD Bに基づいて、第1年齢と同じ第2年齢の第2訓練画像T I 2を取得する(S 4)。

10

**【0 1 0 0】**

学習装置2 0は、現状の学習モデルMと、第2訓練画像T I 2と、に基づいて、第2特徴量F 2と、第2分布D 2と、を取得する(S 5)。S 5の処理は、第2訓練画像T I 2が学習モデルMに入力される点でS 2の処理と異なるが、他の点はS 2の処理と同様である。学習装置2 0は、第2分布D 2に基づいて、第2平均年齢A A 2を取得する(S 6)。第2平均年齢A A 2の計算方法は、先述した通りである。

**【0 1 0 1】**

学習装置2 0は、訓練データベースD Bに基づいて、第1年齢と異なる第3年齢の第3訓練画像T I 3を取得する(S 7)。学習装置2 0は、現状の学習モデルMと、第3訓練画像T I 3と、に基づいて、第3分布D 3を取得する(S 8)。S 8では、学習装置2 0は、学習モデルMに第3訓練画像T I 3を入力し、学習モデルMから出力された第3分布D 3を取得する。学習装置2 0は、第3分布D 3に基づいて、第3平均年齢A A 3を取得する(S 9)。第3平均年齢A A 3の計算方法は、先述した通りである。

20

**【0 1 0 2】**

学習装置2 0は、第1特徴量F 1と、第2特徴量F 2と、に基づいて、コサイン類似度損失を計算する(S 1 0)。学習装置2 0は、第1平均年齢A A 1、第2平均年齢A A 2、及び第3平均年齢A A 3に基づいて、トリプレットマージン損失を計算する(S 1 1)。学習装置2 0は、ソフトマックス損失、平均損失、分散損失、コサイン類似度損失、及びトリプレットマージン損失に基づいて、学習処理を実行する(S 1 2)。

**【0 1 0 3】**

学習装置2 0は、学習処理を終了するか否かを判定する(S 1 3)。学習処理は、任意のタイミングで終了可能である。例えば、訓練データベースD Bの全ての訓練データを学習させた場合に終了してもよいし、所定数の訓練データを学習させた場合に終了してもよい。学習処理を終了すると判定されない場合(S 1 3 ; N)、S 1の処理に戻る。学習処理を終了すると判定された場合(S 1 3 ; Y)、学習装置2 0は、推定装置1 0に対し、学習済みの学習モデルMを送信し(S 1 4)、本処理は終了する。推定装置1 0は、学習済みの学習モデルMを記憶部1 2に記録し、実運用を開始する。

30

**【0 1 0 4】**

本実施形態の学習システム1は、第1訓練画像T I 1と、第2訓練画像T I 2と、を利用した距離学習に基づいて、推定対象画像E Iに含まれる推定対象人間E Hの推定年齢を推定する学習モデルMの学習処理を実行する。距離学習を利用することによって、訓練データが少なかったとしても効率の良い学習処理が可能になるので、学習モデルMの精度が高まる。学習モデルMの作成者が多数の訓練データを用意しなくても済むので、作成者の手間を軽減できる。

40

**【0 1 0 5】**

また、学習システム1は、第1訓練画像T I 1、第2訓練画像T I 2、及び第3訓練画像T I 3を利用した距離学習に基づいて、学習処理を実行する。2枚ではなく3枚の訓練画像T Iを利用することによって、より効率の良い学習処理が可能になるので、学習モデルMの精度がより高まる。より少ない訓練データで精度の高い学習モデルMを作成できるので、作成者の手間をより軽減できる。

50

## 【 0 1 0 6 】

また、学習システム 1 は、第 1 年齢が、第 2 数値と同じであり、かつ、第 3 数値とは異なる場合に、第 1 処理結果と第 2 処理結果との違いが小さくなり、かつ、第 1 処理結果と第 3 処理結果との違いが大きくなるように、学習処理を実行する。例えば、学習システム 1 は、第 1 処理結果 ~ 第 3 処理結果の一例である第 1 平均年齢  $A A 1$  ~ 第 3 平均年齢  $A A 3$  を利用したトリプレットマージン損失に基づいて、学習処理を実行する。これにより、いわゆるアンカー画像とポジティブ画像の関係と、アンカー画像とネガティブ画像との関係と、の両方を利用して学習処理を実行できるので、学習モデル M の精度がより高まる。即ち、似た訓練画像 T I 同士が近づくようにだけでなく、異なる訓練画像 T I 同士が遠ざかるような学習処理が実行されるので、学習モデル M の精度がより高まる。

10

## 【 0 1 0 7 】

また、学習システム 1 は、第 1 処理結果と第 3 処理結果との違いが、第 1 年齢と第 3 年齢との差に応じた違いになるように、学習処理を実行する。例えば、第 1 年齢と第 3 年齢との差に応じた値になるように、トリプレットマージン損失におけるマージン が決定される。これにより、異なるラベルが付与された画像同士が単に遠ざかるように学習処理が実行される従来の距離学習に比べて、年齢差に応じた違いになるように学習処理が実行されるので、学習モデル M の精度がより高まる。即ち、推定年齢を推定するといった学習モデル M の性質を利用して、年齢差に応じてどの程度遠ざけるかを調整することによって、最適な学習処理を実行できるので、学習モデル M の精度がより高まる。

## 【 0 1 0 8 】

また、学習システム 1 は、第 1 推定結果、第 2 推定結果、及び第 3 推定結果に基づいて、学習処理を実行する。例えば、学習システム 1 は、第 1 推定結果 ~ 第 3 推定結果の一例である第 1 平均年齢  $A A 1$  ~ 第 3 平均年齢  $A A 3$  を利用したトリプレットマージン損失に基づいて、学習処理を実行する。これにより、3 枚の訓練画像 T I の互いの関係を考慮して最適な学習処理を実行できるので、学習モデル M の精度がより高まる。

20

## 【 0 1 0 9 】

また、学習システム 1 は、第 1 分布  $D 1$ 、第 2 分布  $D 2$ 、及び第 3 分布  $D 3$  に基づいて、学習処理を実行する。例えば、学習システム 1 は、学習モデル M が推定可能な 0 歳 ~ 100 歳までの個々の年齢  $j$  に応じた確率  $p_j$  を含む第 1 分布  $D 1$  ~ 第 3 分布  $D 3$  に基づいて、学習処理を実行する。これにより、現状の学習モデル M の推定結果をより最適に考慮して学習処理を実行できるので、学習モデル M の精度がより高まる。

30

## 【 0 1 1 0 】

また、学習システム 1 は、第 1 特徴量  $F 1$  と、第 2 特徴量  $F 2$  と、に基づいて、学習処理を実行する。これにより、学習モデル M の中間的な計算結果を学習処理に利用できるため、学習モデル M の精度がより高まる。

## 【 0 1 1 1 】

また、学習システム 1 は、第 1 特徴量  $F 1$  と、第 2 特徴量  $F 2$  と、に基づいて、コサイン類似度を計算し、当該コサイン類似度に基づいて、学習処理を実行する。これにより、現状の学習モデル M の精度をより正確に評価できるコサイン類似度を利用することができるので、学習モデル M の精度がより高まる。

40

## 【 0 1 1 2 】

また、学習システム 1 は、第 1 損失と、関係損失と、に基づいて、学習処理を実行する。例えば、学習システム 1 は、第 1 損失に相当するソフトマックス損失、平均損失、及び分散損失と、関係損失に相当するコサイン類似度損失及びトリプレットマージン損失と、に基づいて、学習処理を実行する。これにより、現状の学習モデル M の精度をより多面的に評価して学習処理を実行できるので、学習モデル M の精度がより高まる。

## 【 0 1 1 3 】

また、学習システム 1 は、複数の第 1 損失と、関係損失と、に基づいて、学習処理を実行する。例えば、学習システム 1 は、単一の第 1 損失ではなく、ソフトマックス損失、平均損失、及び分散損失といった複数の第 1 損失に基づいて、学習処理を実行する。これに

50

より、現状の学習モデルMの精度をより多面的に評価して学習処理を実行できるので、学習モデルMの精度がより高まる。

【0114】

また、本実施形態では、第1オブジェクト及び第2オブジェクトは、互いに異なる人間である。第1数値は、第1オブジェクトの年齢である。第2数値は、第2オブジェクトの年齢である。推定対象オブジェクトは、年齢を推定する対象となる人間である。推定対象数値は、推定対象オブジェクトの年齢である。これにより、推定年齢を推定する学習モデルMの精度が高まる。学習モデルMが推定対象人間EHの推定年齢を精度良く推定できるようになる。その結果、学習モデルMを利用する目的を達成しやすくなる。例えば、推定対象人間EHが店舗の顧客である場合には、店舗の客層をより正確に把握できたり、酒やたばこといった商品を購入する時の年齢確認をより正確に行ったり、年齢に応じた最適な広告を提示したりできるようになる。例えば、空港又はイベント会場といった施設で本人確認をする目的で、推定対象人間EHの年齢が推定される場合には、より正確な本人確認が可能になる。

10

【0115】

[5. 変形例]

なお、本開示は、以上に説明した実施形態に限定されるものではない。本開示の趣旨を逸脱しない範囲で、適宜変更可能である。

【0116】

[5-1. 変形例1]

例えば、実施形態では、第1年齢が第2年齢と同じである場合を例に挙げたが、第1年齢は、第2年齢と異なってもよい。変形例1では、第1年齢が第2年齢及び第3年齢の両方と異なる場合を例に挙げる。ただし、第2年齢は、第3年齢と異なるものとする。第1年齢と第3年齢の差は、第1年齢と第2年齢の差よりも大きいものとする。変形例1では、図3の例において、第2年齢が37歳であるものとする。第1年齢及び第3年齢は、図3と同様に、それぞれ35歳、45歳であるものとする。

20

【0117】

実施形態では、第1年齢と第2年齢が同じだったので、第1処理結果と第2処理結果が互いに近づくように学習処理が実行された。変形例1の場合、第1年齢と第2年齢が異なるので、実施形態とは異なる学習処理が実行される。学習部204は、第1処理結果と第3処理結果との違いが、第1処理結果と第2処理結果との違いよりも大きくなるように、学習処理を実行する。

30

【0118】

例えば、学習部204は、第1処理結果と第2処理結果との違いが、第1年齢と第2年齢との違いに応じた違いになるように、学習処理を実行してもよい。学習部204は、第1処理結果と第3処理結果との違いが、第1年齢と第3年齢との違いに応じた違いになるように、学習処理を実行する。第1年齢と第3年齢との違いは、第1年齢と第2年齢よりも大きいので、第1年齢と第3年齢との違いに応じた違いは、第1年齢と第2年齢との違いに応じた違いよりも大きい。

【0119】

変形例1のように、第1年齢～第3年齢が、それぞれ35歳、37歳、45歳だったとすると、学習部204は、第1処理結果と第2処理結果との違いが、第1年齢と第2年齢との差である2歳に応じた違いになるように、学習処理を実行する。学習部204は、第1処理結果と第3処理結果との違いが、第1年齢と第3年齢との差である10歳に応じた違いになるように、学習処理を実行する。例えば、学習部204は、実施形態と同様にして、トリプレットマージン損失を計算してもよい。ただし、第1年齢と第2年齢が同じではないので、マージンは、実施形態よりも小さいものとする。

40

【0120】

変形例1の学習システム1は、第1年齢と第3年齢の差が、第1年齢と第2年齢の差よりも大きい場合に、第1処理結果と第3処理結果との違いが、第1処理結果と第2処理結

50

果との違いよりも大きくなるように、学習処理を実行する。これにより、訓練データをより有効に活用して学習処理を実行できるので、学習モデルMの精度がより高まる。より少ない訓練データを用意すれば済むので、学習モデルMの作成者の負担が軽減する。

【0121】

[5-2. 変形例2]

例えば、実施形態のトリプレットマージン損失は、第2処理結果と第3処理結果との関係を直接的には考慮されなかったが、学習部204は、第2処理結果と第3処理結果との違いが大きくなるように、学習処理を実行してもよい。図7の例であれば、学習部204は、違い $d_p$ 、 $d_n$ だけではなく、第2平均年齢 $A A_2$ と第3平均年齢 $A A_3$ との違いにも基づいて、学習処理を実行してもよい。

10

【0122】

図9は、変形例2におけるトリプレットマージン損失の一例を示す図である。図9の例では、実施形態と同様に、第1年齢が第2年齢と同じものとする。例えば、学習部204は、第1平均年齢 $A A_1$ と第3平均年齢 $A A_3$ との違い $d_{n1}$ と、第2平均年齢 $A A_2$ と第3平均年齢 $A A_3$ との違い $d_{n2}$ と、の平均値を計算する。学習部204は、実施形態で説明した違い $d_n$ に代えて、この平均値を利用してトリプレットマージン損失を計算する。

【0123】

図9のように、学習部204は、平均値を計算する計算式以外の他の計算式に、第1平均年齢 $A A_1$ と第2平均年齢 $A A_2$ との違い $d_p$ 、第1平均年齢 $A A_1$ と第3平均年齢 $A A_3$ との違い $d_{n1}$ 、及び第2平均年齢 $A A_2$ と第3平均年齢 $A A_3$ との違い $d_{n2}$ を代入し、トリプレットマージン損失を計算してもよい。例えば、学習部204は、違い $d_p$ と $d_{n1}$ との差と、違い $d_p$ と違い $d_{n2}$ との差と、を加算した値に基づいて、トリプレットマージン損失を計算してもよい。

20

【0124】

なお、変形例1及び変形例2を組み合わせた場合も、学習部204は、3つの違いを利用して、トリプレットマージン損失を計算すればよい。例えば、学習部204は、第1平均年齢 $A A_1$ と第3平均年齢 $A A_3$ との違い $d_{n1}$ と、第2平均年齢 $A A_2$ と第3平均年齢 $A A_3$ との違い $d_{n2}$ と、の平均値を計算する。学習部204は、実施形態で説明した違い $d_n$ に代えて、この平均値を利用してトリプレットマージン損失を計算してもよい。この場合も、平均値を計算する計算式以外の他の計算式に基づいて、トリプレットマージン損失が計算されてもよい。

30

【0125】

変形例2の学習システム1は、第2処理結果と第3処理結果との違いが大きくなるように、学習処理を実行する。これにより、訓練データをより有効に活用して学習処理を実行できるので、学習モデルMの精度がより高まる。より少ない訓練データを用意すれば済むので、学習モデルMの作成者の負担が軽減する。

【0126】

[5-3. 変形例3]

例えば、変形例2において、学習部204は、第2処理結果と第3処理結果との違いが、第2年齢と第3年齢の差に応じた違いになるように、学習処理を実行してもよい。実施形態の例であれば、第2年齢と第3年齢の違いが10歳なので、学習部204は、第2平均損失 $A A_2$ と第3平均損失 $A A_3$ との違いが10歳程度の違いになるように、学習処理を実行する。変形例1の例であれば、第2年齢と第3年齢の違いが8歳なので、学習部204は、第2平均損失 $A A_2$ と第3平均損失 $A A_3$ との違いが8歳程度の違いになるように、学習処理を実行する。

40

【0127】

変形例3の学習システム1は、第2処理結果と第3処理結果との違いが、第2年齢と第3年齢の差に応じた違いになるように、学習処理を実行する。例えば、第2年齢と第3年齢との差に応じた値になるように、トリプレットマージン損失におけるマージンが決定

50

される。これにより、異なるラベルが付与された画像同士が遠ざかるように学習処理が実行される従来の距離学習に比べて、年齢差に応じた違いになるように学習処理が実行されるので、学習モデルMの精度がより高まる。即ち、推定年齢を推定するといった学習モデルMの性質を利用した最適な学習処理を実行できるので、学習モデルMの精度がより高まる。

【0128】

[5-4. 変形例4]

例えば、実施形態では、第1推定結果、第2推定結果、及び第3推定結果を利用したトリプレットマージン損失に基づいて学習処理が実行される場合を説明した。第3推定結果が利用されずに、学習部204は、第1推定結果と、第2推定結果と、に基づいて、学習処理を実行してもよい。

10

【0129】

例えば、学習部204は、第1分布D1と、第2分布D2と、に基づいて、カルバックライブラー情報量を計算する。カルバックライブラー情報量は、複数の確率分布Dの違いを評価する指標である。カルバックライブラー情報量の計算式自体は、種々の計算式を利用可能である。学習部204は、カルバックライブラー情報量に基づいて、学習処理を実行する。実施形態のように、第1年齢が第2年齢と同じである場合には、学習部204は、カルバックライブラー情報量が示す違いが小さくなるように、学習処理を実行する。変形例1のように、第1年齢が第2年齢と異なる場合には、学習部204は、カルバックライブラー情報量が示す違いが大きくなるように、学習処理を実行する。

20

【0130】

なお、学習部204は、カルバックライブラー情報量以外の手法を利用して、第1推定結果と、第2推定結果と、に基づいて、学習処理を実行してもよい。例えば、実施形態のように、第1年齢が第2年齢と同じである場合には、学習部204は、第1平均年齢AA1と、第2平均年齢AA2と、が近づくように、学習処理を実行してもよい。変形例1のように、第1年齢が第2年齢と異なる場合には、学習部204は、第1平均年齢AA1と、第2平均年齢AA2と、が遠ざかるように、学習処理を実行してもよい。

【0131】

変形例4の学習システム1は、第1推定結果と、第2推定結果と、に基づいて、学習処理を実行する。これにより、より少ない推定結果で学習処理が完了するので、学習モデルMの精度を高めつつ、学習完了に要する時間を少なくできる。学習装置20は、第3推定結果を取得するための計算を実行しないので、学習装置20の処理負荷を軽減できる。

30

【0132】

また、学習システム1は、第1分布D1と、第2分布D2と、に基づいて、カルバックライブラー情報量を計算し、カルバックライブラー情報量に基づいて、学習処理を実行する。これにより、より信頼性の高い指標を利用して学習処理を実行できるので、学習モデルMの精度がより高まる。

【0133】

[5-5. 変形例5]

例えば、実施形態では、学習部204は、第1損失に相当するソフトマックス損失、平均損失、及び分散損失と、関係損失に相当するコサイン類似度損失及びトリプレットマージン損失と、の合計値が小さくなるように、学習処理を実行した。学習部204は、第1損失、関係損失に関する重み係数、及び関係損失に基づいて、学習処理を実行してもよい。この場合、関係損失が第1損失よりも重要視されるように、重み係数が1よりも大きくなってよい。学習部204は、重み係数を考慮した最終的な損失を計算し、当該最終的な損失が小さくなるように、学習処理を実行してもよい。

40

【0134】

変形例5の学習システム1は、第1損失、関係損失に関する重み係数、及び関係損失に基づいて、学習処理を実行する。これにより、例えば、関係損失が第1損失よりも重要視されるようにすることができるので、学習モデルMの精度が高まる。関係損失だけで学習

50

処理を実行してもよいが、第1損失も重要な指標であることに変わりはないので、第1損失と関係損失をバランス良く考慮して学習処理を実行できる。

【0135】

[5-6. 変形例6]

例えば、実施形態では、関係損失の一例として、コサイン類似度損失及びトリプレットマージン損失を説明した。変形例4で説明したカルバックライブラー情報量も、複数の処理結果の関係を示すものなので、関係損失の一例である。実施形態及び変形例4で説明したように、学習部204は、第1処理結果と、第2処理結果と、に基づいて、複数の関係損失を計算してもよい。学習部204は、第1損失と、複数の関係損失と、に基づいて、学習処理を実行する。関係損失としては、他の損失が利用されてもよい。学習部204は、第1損失と、複数の関係損失と、に基づいて、合計損失を計算し、当該合計損失に基づいて、学習処理を実行する。

10

【0136】

変形例6の学習システム1は、第1損失と、複数の関係損失と、に基づいて、学習処理を実行する。これにより、より多くの関係損失を考慮して学習処理を実行できるので、学習モデルMの精度がより高まる。

【0137】

[5-7. その他変形例]

例えば、上記変形例を組み合わせてもよい。

【0138】

例えば、第1年齢が第2年齢及び第3年齢とは異なるが、第2年齢及び第3年齢が同じであってもよい。この場合、学習部204は、第1処理結果と第2処理結果との違いが大きくなり、かつ、第1処理結果と第3処理結果との違いが大きくなるように、学習処理を実行すればよい。更に、第1年齢、第2年齢、及び第3年齢が全て同じであってもよい。この場合、学習部204は、第1処理結果と第2処理結果との違いが小さくなり、かつ、第1処理結果と第3処理結果との違いが小さくなるように、学習処理を実行すればよい。

20

【0139】

例えば、学習装置20で実現されるものとして説明した機能は、他のコンピュータで実現されてもよいし、複数のコンピュータで分担されてもよい。学習装置20に記憶されるものとして説明したデータは、他のコンピュータ又は情報記憶媒体に記憶されていてもよい。

30

40

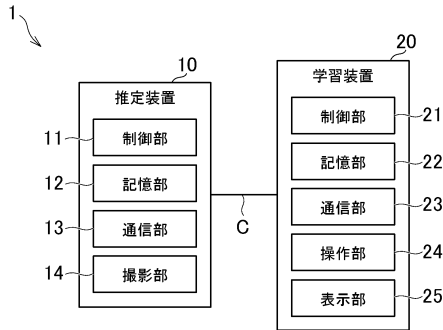
50

【要約】

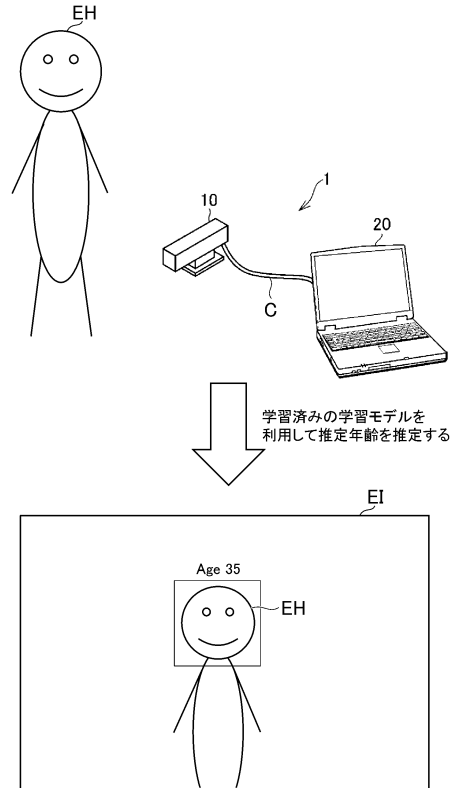
学習システム（S）の第1取得部（201）は、第1数値の第1オブジェクトに関する第1訓練画像を取得する。第2取得部（202）は、第2数値の第2オブジェクトに関する第2訓練画像を取得する。学習部（204）は、第1訓練画像及び第2訓練画像を利用した距離学習に基づいて、推定対象画像に含まれる推定対象オブジェクトに関する推定対象数値を推定する学習モデルの学習処理を実行する。

【図面】

【図1】



【図2】



10

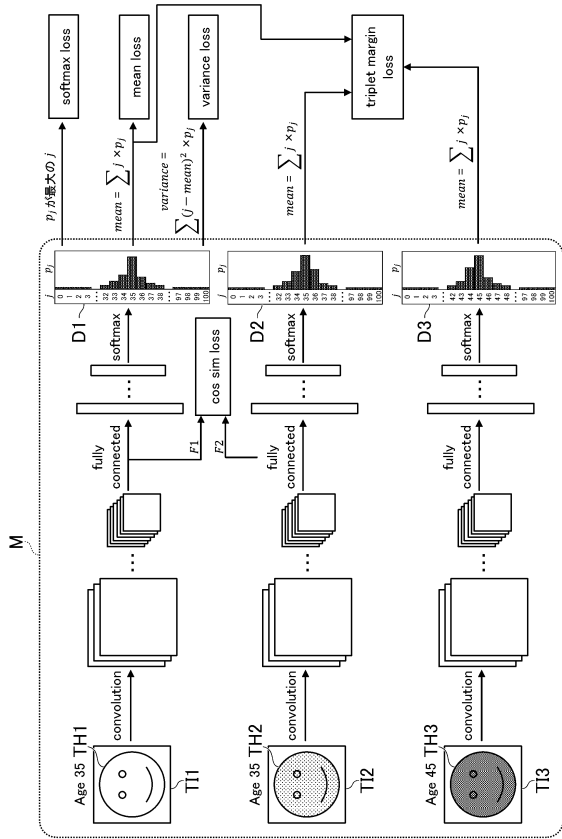
20

30

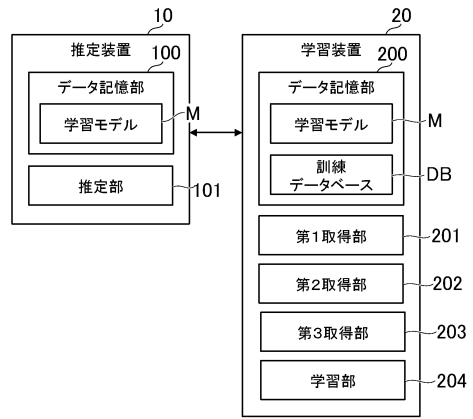
40

50

【図3】



【図4】



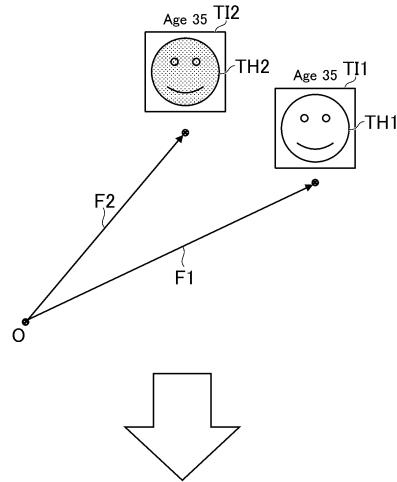
10

20

【図5】

DB	
訓練画像	年齢
00001.jpg	35
00002.jpg	35
00003.jpg	45
...	...
...	...

【図6】



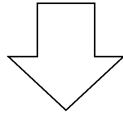
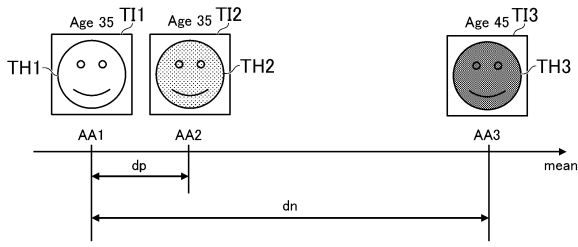
30

40

コサイン類似度損失  $L_{\text{cos sim}} = 1 - \cos(\vec{F1}, \vec{F2})$

50

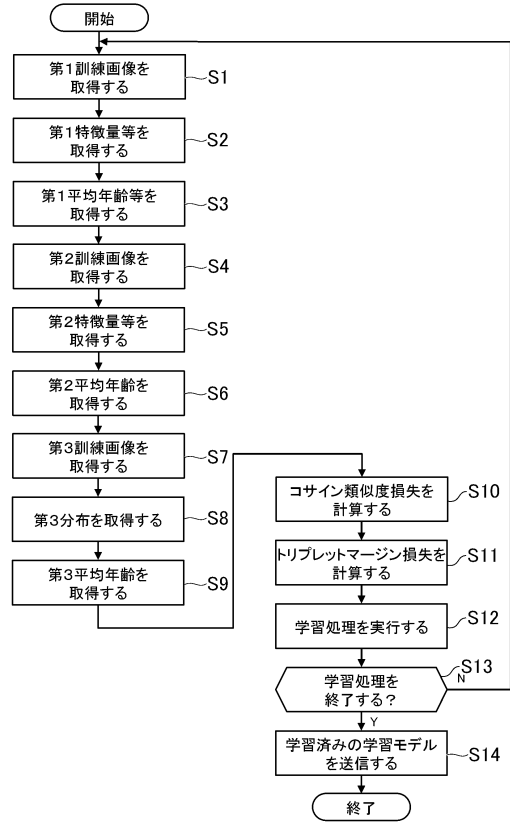
【図7】



トリプレットマージン損失  $L_{triplet} = [dp - dn + \alpha]$

理想的には  $dp + \alpha \leq dn$

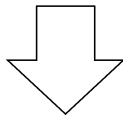
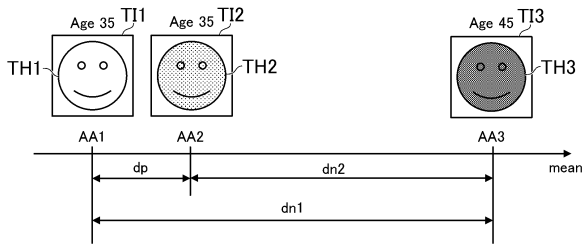
【図8】



10

20

【図9】



トリプレットマージン損失  $L_{triplet} = \left[ dp - \left( \frac{dn1 + dn2}{2} \right) + \alpha \right]$

理想的には  $dp + \alpha \leq \left( \frac{dn1 + dn2}{2} \right)$

30

40

50

## フロントページの続き

東京都世田谷区玉川一丁目14番1号 楽天グループ株式会社内

(72)発明者 シュテンガー ビヨン

東京都世田谷区玉川一丁目14番1号 楽天グループ株式会社内

審査官 福西 章人

(56)参考文献

LIU, H., et al. , Label-Sensitive Deep Metric Learning for Facial Age Estimation , IEEE Transactions on Information Forensics and Security [online] , 2017年08月29日 , Vol. 13, No. 2 , pp. 292-305 , [retrieved on 2022-06-07], Retrieved from the Internet: URL: <https://ieeexplore.ieee.org/abstract/document/8017500> , DOI: 10.1109/TIFS.2017.2746062

Fadi Boutrosa et al. , Self-restrained Triplet Loss for Accurate Masked Face Recognition , arXiv [オンライン] , 2021年12月13日 , インターネット: URL:<https://arxiv.org/pdf/2103.01716v3.pdf>

GAO, B.-B., et al. , Age Estimation Using Expectation of Label Distribution Learning , Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence [online] , 2018年07月13日 , pp. 712-718 , [retrieved on 2022-06-07], Retrieved from the Internet: URL: <https://www.ijcai.org/proceedings/2018/99> , DOI: 10.24963/ijcai.2018/99

(58)調査した分野 (Int.Cl. , D B 名)

G 0 6 N 3 / 0 0 - 9 9 / 0 0

G 0 6 T 7 / 0 0 - 7 / 9 0