

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2015-166975  
(P2015-166975A)

(43) 公開日 平成27年9月24日(2015.9.24)

(51) Int.Cl.	F I	テーマコード (参考)
<b>G06F 17/21 (2006.01)</b>	G06F 17/21 550A	5B091
<b>G06F 17/27 (2006.01)</b>	G06F 17/21 570R	5B109
<b>G06F 17/30 (2006.01)</b>	G06F 17/27 Z	
	G06F 17/30 220C	

審査請求 未請求 請求項の数 6 O L (全 13 頁)

(21) 出願番号 特願2014-41519 (P2014-41519)  
(22) 出願日 平成26年3月4日 (2014.3.4)

(71) 出願人 000005496  
富士ゼロックス株式会社  
東京都港区赤坂九丁目7番3号  
(74) 代理人 100071526  
弁理士 平田 忠雄  
(74) 代理人 100124246  
弁理士 遠藤 和光  
(72) 発明者 榊 茂之  
神奈川県横浜市西区みなとみらい六丁目1番 富士ゼロックス株式会社内  
(72) 発明者 三浦 康秀  
神奈川県横浜市西区みなとみらい六丁目1番 富士ゼロックス株式会社内

最終頁に続く

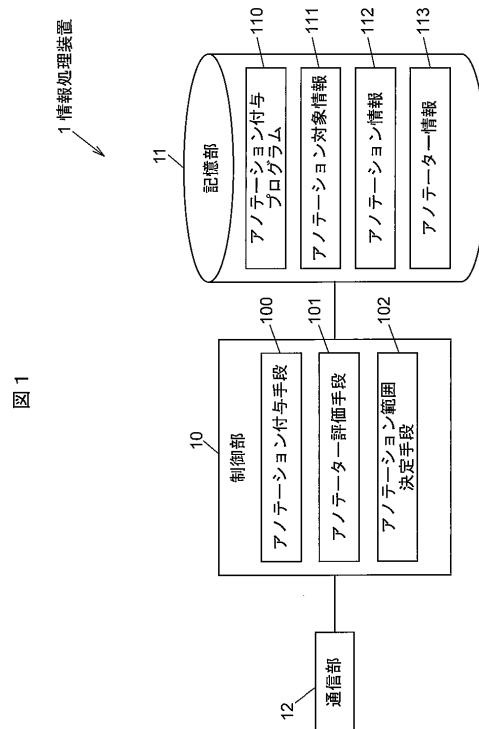
(54) 【発明の名称】 注釈情報付与プログラム及び情報処理装置

(57) 【要約】

【課題】複数の対象となる情報に付与されるアノテーションの信頼性を均一にする注釈情報付与プログラム及び情報処理装置を提供する。

【解決手段】情報処理装置1は、一のアノテーターの入力に基づき、アノテーション対象情報111のアノテーション対象にアノテーションを付与するアノテーション付与手段100と、他のアノテーターによりアノテーション対象に過去に付与されたアノテーションと、一のアノテーターにより付与されたアノテーションとを比較し、一のアノテーター及び他のアノテーターの信頼性を評価するアノテーター評価手段101と、一のアノテーター及び他のアノテーターの信頼性に基づいて、一のアノテーターにアノテーションの付与を依頼するアノテーション対象情報111を決定するアノテーション範囲決定手段102とを有する。

【選択図】 図1



**【特許請求の範囲】****【請求項 1】**

コンピュータを、

一の入力者の入力に基づき、複数の対象を含む対象情報に注釈情報を付与する付与手段と、

他の入力者により前記複数の対象の一部に付与済みの注釈情報と、前記一の入力者により付与された注釈情報とを比較し、前記一の入力者及び前記他の入力者の信頼性を評価する評価手段と、

前記一の入力者及び前記他の入力者の信頼性に基づいて、前記一の入力者に注釈情報の付与を依頼する前記対象情報の対象の範囲を決定する決定手段として機能させるための注釈情報付与プログラム。

10

**【請求項 2】**

前記決定手段は、前記他の入力者の信頼度が予め定めた閾値以上である場合、当該他の入力者が注釈情報を付与した対象以外を、前記一の入力者に注釈情報の付与を依頼する前記対象情報の対象の範囲に決定する請求項 1 に記載の注釈情報付与プログラム。

**【請求項 3】**

前記決定手段は、複数の前記他の入力者の信頼度が予め定めた第 1 の閾値より低いが予め定めた第 2 の閾値以上である場合、複数の当該他の入力者が注釈情報を付与した対象以外を、前記一の入力者に注釈情報の付与を依頼する前記対象情報の対象の範囲に決定する請求項 1 に記載の注釈情報付与プログラム。

20

**【請求項 4】**

機械学習用の情報として、少なくとも、前記対象情報の対象と、前記付与手段によって付与された注釈情報と、当該注釈情報を付与した入力者の信頼性とを有する情報を生成する生成手段としてさらに機能させる請求項 1 - 3 のいずれか 1 項に記載の注釈情報付与プログラム。

**【請求項 5】**

前記生成手段の生成する情報を用いて機械学習を行う機械学習手段としてさらに機能させる請求項 4 に記載の注釈情報付与プログラム。

**【請求項 6】**

一の入力者の入力に基づき、複数の対象を含む対象情報に注釈情報を付与する付与手段と、

30

他の入力者により前記複数の対象の一部に付与済みの注釈情報と、前記一の入力者により付与された注釈情報とを比較し、前記一の入力者及び前記他の入力者の信頼性を評価する評価手段と、

前記一の入力者及び前記他の入力者の信頼性に基づいて、前記一の入力者に注釈情報の付与を依頼する前記対象情報の対象の範囲を決定する決定手段とを有する情報処理装置。

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は、注釈情報付与プログラム及び情報処理装置に関する。

40

**【背景技術】****【0002】**

従来技術として、ヒューマンコンピューテーションの品質を管理する方法が提案されている（例えば、非特許文献 1 参照）。

**【0003】**

非特許文献 1 に開示された方法は、クラウドソーシングを用いて複数の人間に仕事を依頼して結果を受け取り、それをさらに大きな計算過程において活用するヒューマンコンピューテーションにおいて、仕事を依頼した人間（以下、「ワーカー」という。）から得られる結果は一般的に信頼できる特定のワーカーに仕事を依頼した場合と異なり不確実性が大きくなるが、能力の高いワーカーを抽出する、タスクの難しさも考慮してワーカーを評価

50

する、ワーカーの自分の作業品質を自己申告させる、ゲームを取り入れワーカーにまじめに働いてもらう、ワーカーの作業工程をモニタリングする、多数決をとる、ワーカーの能力で票に重みを付ける等の手法を採用することで、個々の人間の仕事の精度を高めたり、タスク選択を効率化する等の手法を採用することで個々のワーカーが間違っただとしても全体としての精度に影響を与えないようにする。

【先行技術文献】

【非特許文献】

【0004】

【非特許文献1】小山 聡著、「ヒューマンコンピューテーションの品質管理」、人工知能、29巻1号、2014年1月、p. 27 - 33

10

【発明の概要】

【発明が解決しようとする課題】

【0005】

本発明の目的は、複数の対象となる情報に付与されるアノテーションの信頼性を均一にする注釈情報付与プログラム及び情報処理装置を提供することにある。

【課題を解決するための手段】

【0006】

本発明の一態様は、上記目的を達成するため、以下の注釈情報付与プログラム及び情報処理装置を提供する。

【0007】

20

[1] コンピュータを、

一の入力者の入力に基づき、複数の対象を含む対象情報に注釈情報を付与する付与手段と、

他の入力者により前記複数の対象の一部に付与済みの注釈情報と、前記一の入力者により付与された注釈情報とを比較し、前記一の入力者及び前記他の入力者の信頼性を評価する評価手段と、

前記一の入力者及び前記他の入力者の信頼性に基づいて、前記一の入力者に注釈情報の付与を依頼する前記対象情報の対象の範囲を決定する決定手段として機能させるための注釈情報付与プログラム。

【0008】

30

[2] 前記決定手段は、前記他の入力者の信頼度が予め定めた閾値以上である場合、当該他の入力者が注釈情報を付与した対象以外を、前記一の入力者に注釈情報の付与を依頼する前記対象情報の対象の範囲に決定する前記[1]に記載の注釈情報付与プログラム。

【0009】

[3] 前記決定手段は、複数の前記他の入力者の信頼度が予め定めた第1の閾値より低いが予め定めた第2の閾値以上である場合、複数の当該他の入力者が注釈情報を付与した対象以外を、前記一の入力者に注釈情報の付与を依頼する前記対象情報の対象の範囲に決定する前記[1]に記載の注釈情報付与プログラム。

【0010】

40

[4] 機械学習用の情報として、少なくとも、前記対象情報の対象と、前記付与手段によって付与された注釈情報と、当該注釈情報を付与した入力者の信頼性とを有する情報を生成する生成手段としてさらに機能させる前記[1] - [3]のいずれかに記載の注釈情報付与プログラム。

【0011】

[5] 前記生成手段の生成する情報を用いて機械学習を行う機械学習手段としてさらに機能させる前記[4]に記載の注釈情報付与プログラム。

【0012】

[6] 一の入力者の入力に基づき、複数の対象を含む対象情報に注釈情報を付与する付与手段と、

他の入力者により前記複数の対象の一部に付与済みの注釈情報と、前記一の入力者によ

50

り付与された注釈情報とを比較し、前記一の入力者及び前記他の入力者の信頼性を評価する評価手段と、

前記一の入力者及び前記他の入力者の信頼性に基づいて、前記一の入力者に注釈情報の付与を依頼する前記対象情報の対象の範囲を決定する決定手段とを有する情報処理装置。

【発明の効果】

【0013】

請求項1又は6に係る発明によれば、複数の対象となる情報に付与される注釈情報の信頼性を均一にすることができる。

【0014】

請求項2に係る発明によれば、複数人に注釈情報の付与を依頼する場合に、信頼性の高い注釈情報が重複して付与されることを抑制することができる。

10

【0015】

請求項3に係る発明によれば、複数人に注釈情報の付与を依頼する場合に、複数の注釈情報を処理することによって信頼性の高い注釈情報が得られるとき、対象に信頼性の高い注釈情報が重複して付与されることを抑制することができる。

【0016】

請求項4に係る発明によれば、機械学習に用いる情報に注釈情報の信頼性を含む情報を生成することができる。

【0017】

請求項5に係る発明によれば、注釈情報の信頼性を考慮して機械学習を実行することができる。

20

【図面の簡単な説明】

【0018】

【図1】図1は、第1の実施の形態に係る情報処理装置の構成例を示すブロック図である。

【図2】図2は、アノテーション対象情報及びアノテーション情報の構成の一例を示す概略図である。

【図3】図3は、アノテーター情報の構成の一例を示す概略図である。

【図4】図4は、アノテーション対象情報とアノテーション情報の構成の一例を示す概略図である。

30

【図5】図5は、情報処理装置の動作の一例を示すフローチャートである。

【図6】図6は、アノテーター情報に付加されるアノテーターメタ情報の構成の一例を示す概略図である。

【図7】図7は、アノテーション対象情報とアノテーション情報の構成の一例を示す概略図である。

【図8】図8は、第2の実施の形態に係る情報処理装置の構成例を示すブロック図である。

【図9】図9は、学習用情報の構成の一例を示す概略図である。

【発明を実施するための形態】

【0019】

40

[第1の実施の形態]

(情報処理装置の構成)

図1は、第1の実施の形態に係る情報処理装置の構成例を示すブロック図である。

【0020】

情報処理装置1は、通信部12を介して外部ネットワークに接続されており、クラウドソーシングにより、外部ネットワーク上に接続された端末等の利用者に、文書情報、画像情報、音声情報等のアノテーション対象情報111に当該情報の特徴等を示す注釈情報であるアノテーションの付与を依頼するものである(以下、アノテーションを付与する利用者を入力者としての「アノテーター」と呼ぶ。)。また、情報処理装置1は、アノテーターからアノテーションの入力を受け付けてアノテーション対象情報111にアノテーショ

50

ンを付与するものである。なお、アノテーションは、「ポジティブ」と「ネガティブ」のように二値のものでもよいし、複数のカテゴリを用意して多値に分類されるものであってもよい。

#### 【0021】

情報処理装置1は、CPU(Central Processing Unit)等から構成され、各部を制御するとともに、各種のプログラムを実行する制御部10と、フラッシュメモリ等の記憶媒体から構成され情報を記憶する記憶部11と、ネットワークを介して外部と通信する通信部12とを備える。

#### 【0022】

制御部10は、後述するアノテーション付与プログラム110を実行することで、アノテーション付与手段100、アノテーター評価手段101及びアノテーション範囲決定手段102等として機能する。

#### 【0023】

アノテーション付与手段100は、アノテーターからアノテーションの入力を受け付けて、複数のアノテーション対象を有するアノテーション対象情報111のうち、いくつかのアノテーション対象にアノテーションを付与する。付与されたアノテーションはアノテーション対象と関連付けられてアノテーション情報112として記憶部11に格納される。

#### 【0024】

アノテーター評価手段101は、同一のアノテーション対象に対し、現在アノテーターが付与しているアノテーションと、他のアノテーターが過去に付与したアノテーションとを比較してアノテーション付与中のアノテーター及び過去にアノテーションを付与したアノテーターの信頼性の評価を行う。評価方法の詳細は後述する。評価結果はアノテーター情報113として記憶部11に格納される。

#### 【0025】

アノテーション範囲決定手段102は、アノテーター評価手段101の評価結果であるアノテーター情報113に基づいてアノテーション付与中のアノテーターに依頼するアノテーション対象情報111中のアノテーション対象の範囲、つまりいずれのアノテーション対象に対してアノテーションの付与を依頼するのかを決定する。範囲の決定方法については後述する。

#### 【0026】

記憶部11は、制御部10を上述した各手段101-102として動作させるアノテーション付与プログラム110、アノテーション対象情報111、アノテーション情報112及びアノテーター情報113等を記憶する。

#### 【0027】

図2は、アノテーション対象情報111及びアノテーション情報112の構成の一例を示す概略図である。

#### 【0028】

アノテーション対象情報111aは、アノテーション対象情報111の一例であり、言語情報に対してアノテーションを行うことを想定した例であり、アノテーション対象として「今日は晴れ。いい天気だ。」といったようなテキストを複数含む文書情報である。

#### 【0029】

アノテーション情報112aは、アノテーション情報112の一例であり、アノテーション対象情報111aのそれぞれのアノテーション対象に対して付与されたアノテーションである。

#### 【0030】

図2に示す例では、アノテーション対象情報111aの各テキストにアノテーションの付与を依頼するアノテーターの数は3人であって、各アノテーターがアノテーションを付与するアノテーション対象の数は3つである。また、付与するアノテーションはポジティブを表す「ポジ」かネガティブを表す「ネガ」のいずれかである。

10

20

30

40

50

## 【 0 0 3 1 】

図 3 は、アノテーター情報 1 1 3 の構成の一例を示す概略図である。

## 【 0 0 3 2 】

アノテーター情報 1 1 3 a は、アノテーター情報 1 1 3 の一例であり、アノテーターを識別するアノテーター欄と、アノテーターの信頼性を示す信頼性欄と、アノテーターがアノテーションを付与したアノテーション対象情報 1 1 1 中のアノテーション対象の範囲を示すアノテーション付与範囲欄とを有する。

## 【 0 0 3 3 】

( 情報処理装置の動作 )

次に、第 1 の実施の形態の作用を、図 1 - 図 5 を参照しつつ、説明する。

10

## 【 0 0 3 4 】

図 4 は、アノテーション対象情報 1 1 1 とアノテーション情報 1 1 2 の構成の一例を示す概略図である。また、図 5 は、情報処理装置の動作の一例を示すフローチャートである。

## 【 0 0 3 5 】

以下に説明する例では、「アノテーター A」及び「アノテーター C」によるアノテーションが既に付与されており、「アノテーター B」にアノテーションの付与を依頼する場合について説明する。また、アノテーション対象情報 1 1 1 b の各アノテーション対象にアノテーションの付与を依頼するアノテーターの数は 3 人であるものとし、各アノテーターは 7 つのアノテーション対象にアノテーションを付与するものとする。

20

## 【 0 0 3 6 】

まず、アノテーション範囲決定手段 1 0 2 は、図 4 に示すアノテーション対象情報 1 1 1 b のうち「教師データ 1」～「教師データ 4」及び「教師データ T + 1」～「教師データ T + 3」の 7 つのアノテーション対象を、アノテーションを付与する範囲 1 0 0 b<sub>1</sub> 及び 1 0 0 b<sub>2</sub> に決定する ( S 1 )。

## 【 0 0 3 7 】

次に、アノテーション付与手段 1 0 0 は、当該範囲 1 0 0 b<sub>1</sub> 及び 1 0 0 b<sub>2</sub> の一部について、例えば、範囲 1 0 0 b<sub>1</sub> の「教師データ 1」～「教師データ 4」について「アノテーター B」にアノテーションの付与を依頼し、「アノテーター B」からアノテーションの入力を受け付けると「教師データ 1」～「教師データ 4」にアノテーションの付与を行う ( S 2 )。この時点でアノテーション情報 1 1 2 b は図 4 に示す状態となる。

30

## 【 0 0 3 8 】

次に、アノテーター評価手段 1 0 1 は、「アノテーター B」が付与している範囲 1 0 0 b<sub>1</sub> のアノテーションと、「アノテーター A」が過去に付与した範囲 1 0 0 a<sub>1</sub> のアノテーション及び「アノテーター C」が過去に付与した範囲 1 0 0 c<sub>1</sub> のアノテーションとを比較して、「アノテーター A」、「アノテーター B」及び「アノテーター C」の信頼性の評価を行う ( S 3 )。

## 【 0 0 3 9 】

図 4 に示す例では、範囲 1 0 0 a<sub>1</sub> のアノテーションと範囲 1 0 0 b<sub>1</sub> のアノテーションとが一致し、範囲 1 0 0 c<sub>1</sub> のアノテーションが「教師データ 3」以外で一致しないため、アノテーター評価手段 1 0 1 はアノテーター情報 1 1 3 a の「アノテーター A」及び「アノテーター B」の信頼性を上げ、「アノテーター C」の信頼性を下げる。この時点で図 3 のアノテーター情報 1 1 3 a に示すように「アノテーター A」及び「アノテーター B」の信頼性が「 8 0 % 」、「アノテーター C」の信頼性が「 5 0 % 」になったものとする。

40

## 【 0 0 4 0 】

次に、アノテーション範囲決定手段 1 0 2 は、アノテーター情報 1 1 3 a を参照し、「アノテーター A」及び「アノテーター B」の信頼性が予め定めた閾値以上である場合、例えば 7 0 % 以上である場合 ( S 4 ; Y e s )、「アノテーター B」に依頼するアノテーション対象情報 1 1 1 b の範囲を、信頼性の高い「アノテーター A」がアノテーションを付

50

与している範囲 100a<sub>2</sub> と重なる範囲 100b<sub>2</sub> を避け、アノテーションを付与していない範囲 100b<sub>3</sub> に設定する (S5)。

【0041】

これは、信頼性の高い「アノテーター A」と同じ範囲に、同様に信頼性の高い「アノテーター B」にアノテーションの付与を依頼しても重複した信頼性の高いアノテーションの付与になる可能性が高いためである。また、同時に信頼性の低い「アノテーター C」と同じ範囲に、信頼性の高い「アノテーター B」にアノテーションの付与を依頼することで信頼性の低いアノテーションの付与が重複することを避けることができるためである。

【0042】

なお、アノテーション付与手段 100 は、「アノテーター A」と「アノテーター B」のように 2 人のアノテーションが一致した場合に信頼性が高いと評価したが、n 人 (n ≥ 3) のアノテーションが一致した場合に信頼性が高いと評価してもよい。

10

【0043】

次に、アノテーション付与手段 100 は、当該範囲 100b<sub>3</sub>、つまり、「教師データ U+1」～「教師データ U+3」について「アノテーター B」にアノテーションの付与を依頼し、「アノテーター B」からアノテーションの入力を受け付けると範囲 100b<sub>3</sub> にアノテーションの付与を行う (S6)。

【0044】

また、ステップ S4 において、アノテーション範囲決定手段 102 は、アノテーター情報 113a を参照し、他のアノテーターの信頼性が予め定めた閾値より小さい場合、例えば 70% より小さい場合 (S4; No)、当初に決定した「教師データ 1」～「教師データ 4」及び「教師データ T+1」～「教師データ T+3」の 7 つのテキストを、アノテーションを付与する範囲のままとする (S7)。

20

【0045】

(第 1 の実施の形態の効果)

上記した第 1 の実施の形態によると、現在入力中のアノテーションと過去に入力されたアノテーションとに基づきアノテーターの信頼性を評価し、信頼性の高いアノテーターが過去にアノテーションを付与している場合、アノテーション対象情報 111 の当該範囲を現在アノテーションを付与中のアノテーターのアノテーション付与範囲から外すようにしたため、複数人にアノテーションの付与を依頼する場合に、信頼性の高いアノテーションが重複して付与されることを抑制することができる。

30

【0046】

[変形例 1]

上記した第 1 の実施の形態のアノテーター情報 113 に、以下に示すメタ情報を追加し、当該情報に基づいてアノテーター評価手段 101 がアノテーターの評価を行うようにしてもよい。

【0047】

図 6 は、アノテーター情報 113 に付加されるアノテーターメタ情報の構成の一例を示す概略図である。

【0048】

アノテーターメタ情報 113A は、アノテーターを識別するアノテーター欄と、アノテーターの性別を示す性別欄と、アノテーターの年齢を示す年齢欄と、アノテーターの国籍を示す国籍欄と、アノテーターの居住地を示す居住地欄とを有する。

40

【0049】

アノテーター評価手段 101 は、例えば、アノテーション対象情報 111 の内容が日本の流行に関する内容であれば、居住地欄が「日本」である「アノテーター A」及び「アノテーター B」が信頼性の高いアノテーションを付与することを前提に、第 1 の実施の形態で説明したようにアノテーションを比較して、その一致又は不一致に基づいて「アノテーター A」及び「アノテーター B」の評価を行うようにしてもよい。

【0050】

50

## [ 変形例 2 ]

上記した第 1 の実施の形態のように、アノテーションの一致又は不一致によりアノテーターを比較する他に、アノテーター評価手段 101 は、以下に示すように単一のアノテーターについて評価を行ってもよい。この方法は、第 1 の実施の形態の評価方法と組み合わせてもよいし、独立して行うものであってもよい。

## 【 0051 】

アノテーター評価手段 101 は、例えば、あるアノテーターが付与したアノテーション情報 112 のエントロピーを計算する。これは、不真面目なアノテーターは単一のアノテーションを全てのデータに付与することがあると考えられるためであり、計算されたエントロピーが小さい場合、信頼性が低いと評価することができる。

10

## 【 0052 】

また、上記した変形例 1、2 の他にも、「アノテーターに自分の作業品質を自己申告させる」、「アノテーターの作業工程をモニタリングする」、「過去の別のアノテーション作業において評価したアノテーターの信頼性を利用する」といったような従来技術を組み合わせ信頼性の評価をしてもよく、これにより評価精度を向上することができるのはもちろんである。

## 【 0053 】

## [ 変形例 3 ]

上記した第 1 の実施の形態において説明したアノテーション範囲決定手段 102 の動作に加え、以下に説明するようにアノテーション範囲決定手段 102 を動作させてもよい。

20

## 【 0054 】

図 7 は、アノテーション対象情報 111 とアノテーション情報 112 の構成の一例を示す概略図である。

## 【 0055 】

アノテーション対象情報 111 c に対しアノテーション情報 112 c が付与されている場合において、範囲 100 e<sub>1</sub>、100 f<sub>1</sub> 及び 100 f<sub>2</sub> のアノテーションがそれぞれ「教師データ 3」、「教師データ 4」及び「教師データ T + 3」のアノテーションとしては誤りであるものとする。

## 【 0056 】

また、「アノテーター D」、「アノテーター E」、「アノテーター F」の信頼性は、予め定めた閾値（70%）より小さく、予め定めた第 2 の閾値（60%）以上の信頼性であるとする。

30

## 【 0057 】

以上の条件において、信頼性の高いアノテーター（70%以上）に比べて信頼性が劣るが、ある程度信頼性が保障されているアノテーター（60%以上）の場合であって、アノテーションが所定の数、例えば、3つ付与されている場合、アノテーション範囲決定手段 102 は、アノテーション情報 112 c の「教師データ 1」～「教師データ T + 3」の範囲には、さらなるアノテーションが不要と判断し、現在アノテーションを付与中のアノテーターに対し他の範囲のアノテーション付与を依頼するようにしてもよい。

## 【 0058 】

40

## [ 第 2 の実施の形態 ]

以下に第 2 の実施の形態に係る情報処理装置 1A について説明する。第 2 の実施の形態は、第 1 の実施の形態と比べて、アノテーション対象情報 111、アノテーション情報 112 及びアノテーター情報 113 に基づき機械学習に用いる情報を生成し、当該情報を用いて機械学習を行う点で異なる。なお、第 1 の実施の形態と共通の構成については同一の符号を付している。

## 【 0059 】

図 8 は、第 2 の実施の形態に係る情報処理装置の構成例を示すブロック図である。

## 【 0060 】

情報処理装置 1A は、第 1 の実施の形態の情報処理装置 1 と比べて、学習用情報生成手

50



段 1 0 3、機械学習手段 1 0 4 及び学習用情報 1 1 4 をさらに有する。

【 0 0 6 1 】

学習用情報生成手段 1 0 3 は、アノテーション対象情報 1 1 1、アノテーション情報 1 1 2 及びアノテーター情報 1 1 3 に基づいて学習用情報 1 1 4 を生成する。

【 0 0 6 2 】

機械学習手段 1 0 4 は、学習用情報 1 1 4 を用いて機械学習を実行する。

【 0 0 6 3 】

図 9 は、学習用情報 1 1 4 の構成の一例を示す概略図である。

【 0 0 6 4 】

学習用情報 1 1 4 a は、学習用情報 1 1 4 の一例であり、アノテーション欄と、アノテーター欄と、信頼性欄と、アノテーション対象情報欄とを有する。 10

【 0 0 6 5 】

( 情報処理装置の動作 )

次に、第 2 の実施の形態の作用を説明する。

【 0 0 6 6 】

情報処理装置 1 A は、各手段 1 0 0 ~ 1 0 2 によりアノテーション対象情報 1 1 1 にアノテーション情報 1 1 2 を付与するとともに、アノテーター情報 1 1 3 を生成する。

【 0 0 6 7 】

次に、学習用情報生成手段 1 0 3 は、アノテーション対象情報 1 1 1 とアノテーション情報 1 1 2 とから構成される一般的な機械学習用の情報に、アノテーター情報 1 1 3 に含まれる項目をさらに追加して学習用情報 1 1 4 とする。図 9 に示す例では、学習用情報 1 1 4 d は、一般的な機械学習用の情報としてアノテーション対象情報 1 1 1 に対応するアノテーション対象情報欄と、アノテーション情報 1 1 2 に対応するアノテーション欄とを有し、さらにアノテーター情報 1 1 3 に含まれるアノテーター欄と、信頼性欄とを有する。 20

【 0 0 6 8 】

次に、機械学習手段 1 0 4 は、学習用情報 1 1 4 a を用いて機械学習を行う。この際に信頼性欄の値を考慮して学習用情報 1 1 4 a のそれぞれについて重み付けを行ってもよい。また、さらにアノテーターメタ情報 1 1 3 A を用いて重み付けを行ってもよい。

【 0 0 6 9 】

( 第 2 の実施の形態の効果 )

上記した第 2 の実施の形態によれば、機械学習用の情報として用いられる情報は通常アノテーション対象とアノテーションのみであるが、当該機械学習用の情報にアノテーターの信頼性を追加したものとしたため、アノテーションの信頼性を考慮した機械学習用の情報を生成することができ、アノテーションの信頼性を考慮した機械学習を実行することができる。 30

【 0 0 7 0 】

[ 他の実施の形態 ]

なお、本発明は、上記実施の形態に限定されず、本発明の趣旨を逸脱しない範囲で種々な変形が可能である。 40

【 0 0 7 1 】

上記実施の形態では制御部 1 0 の各手段 1 0 0 ~ 1 0 4 の機能をプログラムで実現したが、各手段の全て又は一部を A S I C 等のハードウェアによって実現してもよい。また、上記実施の形態で用いたプログラムを C D - R O M 等の記録媒体に記憶して提供することもできる。また、上記実施の形態で説明した上記ステップの入れ替え、削除、追加等は本発明の要旨を変更しない範囲内で可能である。

【 符号の説明 】

【 0 0 7 2 】

1、1 A 情報処理装置

1 0 制御部 50



【 図 3 】

図 3

113a

アノテーター	信頼性	アノテーション付与範囲
A	80%	1 ~ 3
B	80%	4 ~ 6
C	50%	1 ~ 3
D	60%	4 ~ 6
⋮	⋮	⋮

【 図 4 】

図 4

アノテーション対象情報 111b	アノテーション情報 112b		
	アノテーター-A	アノテーター-B	アノテーター-C
教師タータ1	X	X	Y
教師タータ2	X	X	Y
教師タータ3	Y	Y	Y
教師タータ4	X	X	Y
⋮	100a1	100b1	100c1
教師タータT+1	X		
教師タータT+2	Y		
教師タータT+3	Y		
⋮	100a2	100b2	100c2
教師タータU+1			Y
教師タータU+2			X
教師タータU+3			X
			100c3

【 図 6 】

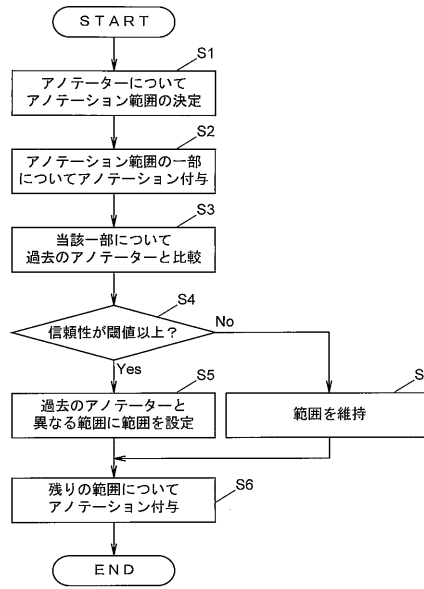
図 6

113A アノテーターメタ情報

アノテーター	性別	年齢	国籍	居住地
A	男	20代	日本	日本
B	男	20代	日本	日本
C	男	40代	日本	アメリカ
D	女	30代	アメリカ	アメリカ
⋮	⋮	⋮	⋮	⋮

【 図 5 】

図 5



【 図 7 】

図 7

アノテーション対象情報 111c	アノテーション情報 112c			
	アノテーター-D	アノテーター-E	アノテーター-F	アノテーター-G
教師タータ1	X	X	X	100e1
教師タータ2	X	X	X	
教師タータ3	Y	X	Y	100f1
教師タータ4	X	X	Y	
⋮	⋮	⋮	⋮	100f2
教師タータT+1	X	X	X	
教師タータT+2	Y	Y	Y	
教師タータT+3	Y	Y	X	
⋮	⋮	⋮	⋮	

【 図 8 】

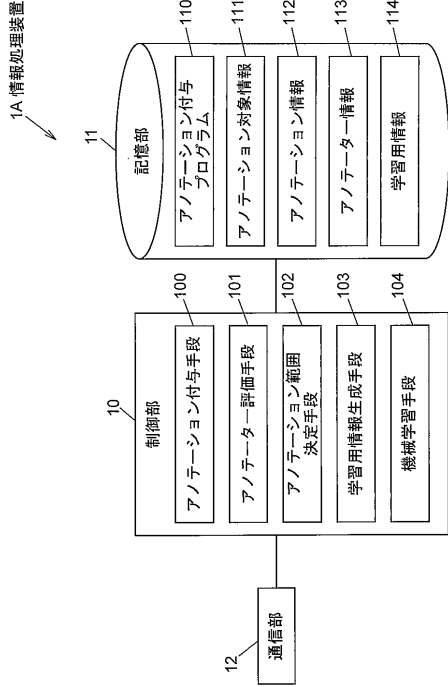


図 8

【 図 9 】

114d 学習用情報

アノテーション	アノテーター	信頼性	アノテーション対象情報
ネガ	A	80%	風邪を引いたかもしれない。
ネガ	B	50%	風邪を引いたかもしれない。
ネガ	C	60%	風邪を引いたかもしれない。
ネガ	F	50%	風邪を引いたかもしれない。
ポジ	D	95%	久しぶりに友人に会った。元気そうだった。

図 9

---

フロントページの続き

- (72)発明者 服部 圭悟  
神奈川県横浜市西区みなとみらい六丁目1番 富士ゼロックス株式会社内
- (72)発明者 坪下 幸寛  
神奈川県横浜市西区みなとみらい六丁目1番 富士ゼロックス株式会社内
- (72)発明者 大熊 智子  
神奈川県横浜市西区みなとみらい六丁目1番 富士ゼロックス株式会社内
- Fターム(参考) 5B091 AA15 CA12  
5B109 QA04 SA14