(19)日本国特許庁(JP)

(12)特許公報(B1)

(11)特許番号
特許第7366204号
(P7366204)

(45)発行日　令和5年10月20日(2023.10.20)

(24)登録日　令和5年10月12日(2023.10.12)

(51)国際特許分類
Ｇ０６Ｆ　16/53　(2019.01)

ＦＩ
Ｇ０６Ｆ　16/53

請求項の数　12　（全21頁）

(21)出願番号　特願2022-116617(P2022-116617)
(22)出願日　令和4年7月21日(2022.7.21)
審査請求日　令和5年4月14日(2023.4.14)
早期審査対象出願

(73)特許権者　517255566
株式会社エクサウィザーズ
東京都港区東新橋一丁目９番２号
(74)代理人　100114557
弁理士　河野 英仁
(74)代理人　100078868
弁理士　河野 登夫
(72)発明者　小島 啓明
東京都港区東新橋一丁目９番２号 株式会社エクサウィザーズ内
(72)発明者　稲葉 正樹
東京都港区東新橋一丁目９番２号 株式会社エクサウィザーズ内
(72)発明者　佐々木 励
東京都港区東新橋一丁目９番２号 株式
最終頁に続く

(54)【発明の名称】　情報処理方法、コンピュータプログラム及び情報処理装置

(57)【特許請求の範囲】
【請求項１】
情報処理装置が、
画像及び当該画像に対応するテキストが対応付けられた正例の組を複数取得し、
画像及びテキストの入力を受け付けて前記画像及び前記テキストの類似度を出力するよう機械学習がなされた学習モデルへ、取得した正例の組の各組の画像及びテキストを入力して前記学習モデルが出力する前記画像及び前記テキストの類似度を取得することで、複数の前記正例の組の類似度をそれぞれ取得し、
取得した類似度の分布に基づいて閾値を決定し、
処理対象となる複数の画像を取得し、
前記複数の画像からの画像の抽出条件となるテキストを取得し、
取得した画像及びテキストを前記学習モデルへ入力して前記学習モデルが出力する前記画像及び前記テキストの類似度を取得することで、前記複数の画像と前記テキストとの各組の類似度を取得し、
取得した各組の類似度と、決定した前記閾値とを比較し、
前記複数の画像から、前記類似度が前記閾値を超える画像を抽出して出力する、
情報処理方法。
【請求項２】
情報処理装置が、
処理対象となる複数の画像を取得し、

前記複数の画像からの画像の抽出条件となるテキストを取得し、
画像及びテキストの入力を受け付けて前記画像及び前記テキストの類似度を出力するよう機械学習がなされた学習モデルへ、取得した画像及びテキストを入力して前記学習モデルが出力する前記画像及び前記テキストの類似度を取得することで、前記複数の画像と前記テキストとの各組の類似度を取得し、
取得した各組の類似度の分布を表示部に表示し、
前記分布に基づいて閾値の設定を受け付け、
取得した各組の類似度と、受け付けた前記閾値とを比較し、
前記複数の画像から、前記類似度が前記閾値を超える画像を抽出して出力する、
情報処理方法。

【請求項３】
前記情報処理装置が、
前記画像及び当該画像に対応するテキストが対応付けられた正例の組と、前記画像及び当該画像に対応しないテキストとが対応付けられた負例の組とをそれぞれ複数取得し、
取得した正例の組及び負例の組の各組ついて前記学習モデルによる類似度をそれぞれ取得し、
取得した類似度に基づいて、前記学習モデルの適合度又は再現度を算出し、
算出した前記適合度又は前記再現度に基づいて前記閾値を決定する、
請求項１又は請求項２に記載の情報処理方法。

【請求項４】
前記情報処理装置が、
取得した各組の類似度の分布に基づいて前記閾値を決定する、
請求項１又は請求項２に記載の情報処理方法。

【請求項５】
前記情報処理装置が、
取得した各組の類似度に基づいて前記複数の画像を順位付けし、
所定の順位の画像を抽出するよう前記閾値を決定する、
請求項４に記載の情報処理方法。

【請求項６】
前記情報処理装置が、
前記類似度の分布に関するパラメータを算出し、
前記類似度の分布が所定分布であるとみなし、算出した前記パラメータに応じて前記閾値を決定する、
請求項４に記載の情報処理方法。

【請求項７】
前記学習モデルは、
入力された画像の特徴量を出力する画像エンコーダと、
入力されたテキストの特徴量を出力するテキストエンコーダと、
前記画像エンコーダが出力した特徴量及び前記テキストエンコーダが出力した特徴量を基に類似度を算出する算出部と
を有する、
請求項１又は請求項２に記載の情報処理方法。

【請求項８】
前記学習モデルは、大規模汎用画像モデルである、
請求項１又は請求項２に記載の情報処理方法。

【請求項９】
コンピュータに、
画像及び当該画像に対応するテキストが対応付けられた正例の組を複数取得し、
画像及びテキストの入力を受け付けて前記画像及び前記テキストの類似度を出力するよう機械学習がなされた学習モデルへ、取得した正例の組の各組の画像及びテキストを入力し

て前記学習モデルが出力する前記画像及び前記テキストの類似度を取得することで、複数の前記正例の組の類似度をそれぞれ取得し、
取得した類似度の分布に基づいて閾値を決定し、
処理対象となる複数の画像を取得し、
前記複数の画像からの画像の抽出条件となるテキストを取得し、
取得した画像及びテキストを前記学習モデルへ入力して前記学習モデルが出力する前記画像及び前記テキストの類似度を取得することで、前記複数の画像と前記テキストとの各組の類似度を取得し、
取得した各組の類似度と、決定した前記閾値とを比較し、
前記複数の画像から、前記類似度が前記閾値を超える画像を抽出して出力する
処理を実行させる、コンピュータプログラム。
【請求項１０】
コンピュータに、
処理対象となる複数の画像を取得し、
前記複数の画像からの画像の抽出条件となるテキストを取得し、
画像及びテキストの入力を受け付けて前記画像及び前記テキストの類似度を出力するよう機械学習がなされた学習モデルへ、取得した画像及びテキストを入力して前記学習モデルが出力する前記画像及び前記テキストの類似度を取得することで、前記複数の画像と前記テキストとの各組の類似度を取得し、
取得した各組の類似度の分布を表示部に表示し、
前記分布に基づいて閾値の設定を受け付け、
取得した各組の類似度と、受け付けた前記閾値とを比較し、
前記複数の画像から、前記類似度が前記閾値を超える画像を抽出して出力する
処理を実行させる、コンピュータプログラム。
【請求項１１】
画像及び当該画像に対応するテキストが対応付けられた正例の組を複数取得する正例取得部と、
画像及びテキストの入力を受け付けて前記画像及び前記テキストの類似度を出力するよう機械学習がなされた学習モデルへ、取得した正例の組の各組の画像及びテキストを入力して前記学習モデルが出力する前記画像及び前記テキストの類似度を取得することで、複数の前記正例の組の類似度をそれぞれ取得する第１類似度取得部と、
取得した類似度の分布に基づいて閾値を決定する閾値決定部と、
処理対象となる複数の画像を取得する画像取得部と、
前記複数の画像からの画像の抽出条件となるテキストを取得するテキスト取得部と、
取得した画像及びテキストを前記学習モデルへ入力して前記学習モデルが出力する前記画像及び前記テキストの類似度を取得することで、前記複数の画像と前記テキストとの各組の類似度を取得する第２類似度取得部と、
取得した各組の類似度と所定の閾値とを比較する比較部と、
前記複数の画像から、前記類似度が前記閾値を超える画像を抽出して出力する出力部と
を備える、情報処理装置。
【請求項１２】
処理対象となる複数の画像を取得する画像取得部と、
前記複数の画像からの画像の抽出条件となるテキストを取得するテキスト取得部と、
画像及びテキストの入力を受け付けて前記画像及び前記テキストの類似度を出力するよう機械学習がなされた学習モデルへ、取得した画像及びテキストを入力して前記学習モデルが出力する前記画像及び前記テキストの類似度を取得することで、前記複数の画像と前記テキストとの各組の類似度を取得する類似度取得部と、
取得した各組の類似度の分布を表示部に表示する表示処理部と、
前記分布に基づいて閾値の設定を受け付ける受付部と、
取得した各組の類似度と、受け付けた前記閾値とを比較する比較部と、

前記複数の画像から、前記類似度が前記閾値を超える画像を抽出して出力する出力部と
を備える、情報処理装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、複数の画像から所望の画像を抽出する処理を行う情報処理方法、コンピュータプログラム及び情報処理装置に関する。
【背景技術】
【０００２】
特許文献１においては、画像群の中から、画像に付与されたタグ情報を使用して画像を検索する画像処理装置が提案されている。この画像処理装置は、画像群に含まれる全ての画像に付与された全てのタグ情報の少なくとも一部を表示部に表示し、表示したタグ情報の中からユーザの指示に応じて選択された１つ目のタグ情報を第１選択タグ情報として指定し、画像群の中から第１選択タグ情報が付与された画像を第１検索画像として抽出する。画像処理装置は、全ての第１検索画像の少なくとも一部を表示部に表示し、全ての第１検索画像に付与された全てのタグ情報の少なくとも一部を表示部に表示する。
【先行技術文献】
【特許文献】
【０００３】
【文献】特開２０２２－６６３４２号公報
【発明の概要】
【発明が解決しようとする課題】
【０００４】
特許文献１に記載の技術では、画像群の中の全ての画像に対してタグ情報が付与されていることを前提として、画像の検索及び抽出等の処理が行われている。このため特許文献１に記載の画像処理装置は、タグ情報が付与されていない画像を検索及び抽出等の対象とすることはできない。また特許文献１に記載の画像処理装置は、タグ情報としていずれの画像にも付与されていない単語又は文言等をキーワードとして画像の検索及び抽出等を行うことはできない。
【０００５】
本発明は、斯かる事情に鑑みてなされたものであって、その目的とするところは、テキストに基づく画像の検索及び抽出等を実現することが期待できる情報処理方法、コンピュータ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】
一実施形態に係る情報処理方法は、情報処理装置が、画像及び当該画像に対応するテキストが対応付けられた正例の組を複数取得し、画像及びテキストの入力を受け付けて前記画像及び前記テキストの類似度を出力するよう機械学習がなされた学習モデルへ、取得した正例の組の各組の画像及びテキストを入力して前記学習モデルが出力する前記画像及び前記テキストの類似度を取得することで、複数の前記正例の組の類似度をそれぞれ取得し、取得した類似度の分布に基づいて閾値を決定し、処理対象となる複数の画像を取得し、前記複数の画像からの画像の抽出条件となるテキストを取得し、取得した画像及びテキストを前記学習モデルへ入力して前記学習モデルが出力する前記画像及び前記テキストの類似度を取得することで、前記複数の画像と前記テキストとの各組の類似度を取得し、取得した各組の類似度と、決定した前記閾値とを比較し、前記複数の画像から、前記類似度が前記閾値を超える画像を抽出して出力する。
また一実施形態に係る情報処理方法は、情報処理装置が、処理対象となる複数の画像を取得し、前記複数の画像からの画像の抽出条件となるテキストを取得し、画像及びテキストの入力を受け付けて前記画像及び前記テキストの類似度を出力するよう機械学習がなされた学習モデルへ、取得した画像及びテキストを入力して前記学習モデルが出力する前記画

像及び前記テキストの類似度を取得することで、前記複数の画像と前記テキストとの各組の類似度を取得し、取得した各組の類似度の分布を表示部に表示し、前記分布に基づいて閾値の設定を受け付け、取得した各組の類似度と、受け付けた前記閾値とを比較し、前記複数の画像から、前記類似度が前記閾値を超える画像を抽出して出力する。

【発明の効果】

【０００７】

一実施形態による場合は、テキストに基づく画像の検索及び抽出等を実現することが期待できる。

【図面の簡単な説明】

【０００８】

【図１】本実施の形態に係る情報処理システムの概要を説明するための模式図である。

【図２】本実施の形態に係るサーバ装置の構成を示すブロック図である。

【図３】本実施の形態に係る情報処理システムが使用する学習モデルの一構成例を示す模式図である。

【図４】学習モデルの学習方法の概要を説明するための模式図である。

【図５】本実施の形態に係るサーバ装置が行う画像抽出処理の手順を示すフローチャートである。

【図６】適合度に基づく閾値の決定方法を説明するための模式図である。

【図７】再現度に基づく閾値の決定方法を説明するための模式図である。

【図８】代表値に基づく閾値の決定方法を説明するための模式図である。

【図９】分布に基づく閾値の決定を説明するための模式図である。

【図１０】ユーザによる閾値の決定方法を説明するための模式図である。

【発明を実施するための形態】

【０００９】

本発明の実施形態に係る情報処理システムの具体例を、以下に図面を参照しつつ説明する。なお、本発明はこれらの例示に限定されるものではなく、特許請求の範囲によって示され、特許請求の範囲と均等の意味及び範囲内でのすべての変更が含まれることが意図される。

【００１０】

＜システム概要＞

図１は、本実施の形態に係る情報処理システムの概要を説明するための模式図である。本実施の形態に係る情報処理システムは、一又は複数のカメラ３が撮影した画像を、サーバ装置１が取得して画像ＤＢ（データベース）５に記憶して蓄積するシステムである。ユーザは例えば端末装置７を利用してサーバ装置１にアクセスし、画像ＤＢ５に蓄積された画像の閲覧及び取得（ダウンロード）等を行うことができる。この際に本実施の形態に係る情報処理システムでは、ユーザがキーワード等のテキストを入力することで、このテキストに応じた画像の検索又は抽出等を行うことが可能である。

【００１１】

カメラ３は、例えば所定の施設に設置されたカメラ、自動車等の移動体に搭載されたカメラ、又は、ユーザが所持するカメラ等の種々のカメラであってよい。ユーザが所持するカメラには、例えばスマートフォン又はタブレット型端末装置等の情報処理装置に搭載されたカメラが含まれ得る。本実施の形態に置いてカメラ３は、例えばインターネット又は携帯電話通信網等のネットワークを介してサーバ装置１との通信を行うことが可能であり、撮影した画像をサーバ装置１へ送信する。カメラ３は、例えば撮影を行う毎に画像をサーバ装置１へ送信してもよく、例えば１時間に１回又は１日に１回等の周期で撮影した画像をまとめてサーバ装置１へ送信してもよく、また例えばユーザによるアップロードの操作に応じてユーザが選択した画像をサーバ装置１へ送信してもよく、これら以外の適宜のタイミングで画像をサーバ装置１へ送信してよい。またカメラ３が通信機能を備えていなくてもよく、この場合には例えばメモリカード等の記録媒体又は通信機能を有するスマートフォン等の端末装置等を介して、サーバ装置１との間で画像の授受が行われてもよい。

【００１２】
なお画像ＤＢ５が記憶する画像には、静止画像のみでなく、動画像が含まれ得る。サーバ装置１は、画像ＤＢ５に記憶された動画像から静止画像を抽出して後続の処理に用いてもよく、カメラ３が撮影した動画像から静止画像を抽出して画像ＤＢ５に記憶してもよい。また画像ＤＢ５が記憶する画像は、カメラ３が撮影した画像に限らず、例えばユーザがインターネット等を介してダウンロードした画像、ユーザが描いた画像、又は、ゲーム等の画面をキャプチャして取得した画像等の種々の画像が含まれてよい。
【００１３】
端末装置７は、例えばパーソナルコンピュータ、スマートフォン又はタブレット型端末装置等の汎用の情報処理装置が用いられ得る。汎用の情報処理装置に本実施の形態に係る情報処理システムが提供するアプリケーションプログラムをインストールするか、又は、汎用のインターネットブラウザのプログラムを利用してサーバ装置１にアクセスすることによって、ユーザは端末装置７を用いて本実施の形態に係る情報処理システムが提供する画像配信のサービスを利用することができる。端末装置７は、インターネット等のネットワークを介してサーバ装置１との通信を行うことができる。端末装置７は、例えばユーザからテキストの入力を受け付けてサーバ装置１へ送信し、これに応じてサーバ装置１が送信する一又は複数の画像を受信し、受信した画像を表示部に表示する。
【００１４】
サーバ装置１は、例えば本実施の形態に係る情報処理システムを提供する企業等が管理及び運営する装置である。サーバ装置１は、例えばクラウドサーバ等の仮想的なサーバ装置であってもよい。本実施の形態に係るサーバ装置１は、画像を記憶して蓄積するための画像ＤＢ５を備えている。サーバ装置１は、インターネット等のネットワークを介した通信を行い、一又は複数のカメラ３が撮影した画像を取得して画像ＤＢ５に記憶する。またサーバ装置１は、端末装置７からの要求に応じて又は所定のタイミングで、画像ＤＢ５から一又は複数の画像を読み出して端末装置７へ送信する。本実施の形態に係るサーバ装置１は、端末装置７がユーザから受け付けたテキストの情報を取得し、画像ＤＢ５に記憶された複数の画像の中からこのテキストに応じた画像を抽出し、抽出した一又は複数の画像を端末装置７へ送信する。
【００１５】
＜装置構成＞
図２は、本実施の形態に係るサーバ装置１の構成を示すブロック図である。本実施の形態に係るサーバ装置１は、処理部１１、記憶部（ストレージ）１２及び通信部（トランシーバ）１３等を備えて構成されている。なお本実施の形態においては、１つのサーバ装置１にて処理が行われるものとして説明を行うが、複数のサーバ装置が分散して処理を行ってもよい。
【００１６】
処理部１１は、ＣＰＵ（Central Processing Unit）、ＭＰＵ（Micro-Processing Unit）、ＧＰＵ（Graphics Processing Unit）又は量子プロセッサ等の演算処理装置、ＲＯＭ（Read Only Memory）及びＲＡＭ（Random Access Memory）等を用いて構成されている。処理部１１は、記憶部１２に記憶されたプログラム１２ａを読み出して実行することにより、カメラ３から画像を取得して画像ＤＢ５に記憶する処理、及び、画像ＤＢ５に記憶した画像の中からテキストに応じた画像を抽出して端末装置７へ送信する処理等の種々の処理を行う。
【００１７】
記憶部１２は、例えばハードディスク又はＳＳＤ（Solid State Drive）等の大容量の記憶装置を用いて構成されている。記憶部１２は、処理部１１が実行する各種のプログラム、及び、処理部１１の処理に必要な各種のデータを記憶する。本実施の形態において記憶部１２は、処理部１１が実行するプログラム１２ａを記憶する。また記憶部１２には、テキストに応じた画像を抽出する処理に用いられる学習済の学習モデルに関する情報を記憶する学習モデル記憶部１２ｂ、及び、カメラ３が撮影した画像を記憶する画像ＤＢ５が

設けられている。

【００１８】

本実施の形態においてプログラム（コンピュータプログラム、プログラム製品）１２ａは、メモリカード又は光ディスク等の記録媒体９９に記録された態様で提供され、サーバ装置１は記録媒体９９からプログラム１２ａを読み出して記憶部１２に記憶する。ただし、プログラム１２ａは、例えばサーバ装置１の製造段階において記憶部１２に書き込まれてもよい。また例えばプログラム１２ａは、遠隔の他のサーバ装置等が配信するものをサーバ装置１が通信にて取得してもよい。例えばプログラム１２ａは、記録媒体９９に記録されたものを書込装置が読み出してサーバ装置１の記憶部１２に書き込んでもよい。プログラム１２ａは、ネットワークを介した配信の態様で提供されてもよく、記録媒体９９に記録された態様で提供されてもよい。

【００１９】

学習モデル記憶部１２ｂは、テキストに応じた画像の抽出処理に用いられる学習済みの学習モデルに関する情報を記憶する。学習モデルに関する情報には、例えば学習モデルがどのような構成であるかを示す構成情報、及び、機械学習の処理により決定された学習モデルの内部パラメータの値等の情報が含まれ得る。本実施の形態においてサーバ装置１は、学習モデルを生成する機械学習の処理を行わず、他の装置が生成した学習済みの学習モデルを取得し、取得した学習モデルを学習モデル記憶部１２ｂに記憶して使用する。ただし、サーバ装置１が機械学習の処理を行って学習モデルを生成してもよい。

【００２０】

本実施の形態に係るサーバ装置１が用いる学習モデルは、画像及びテキストの入力を受け付けて、画像及びテキストの類似度を出力するように予め機械学習がなされた学習モデルである。本実施の形態に係る学習モデルには、例えば大規模汎用画像モデルであるＣＬＩＰ（Contrastive Language-Image Pre-training）の学習モデルが採用され得る。なお本実施の形態においては、学習モデルとしてＣＬＩＰを採用するが、学習モデルはＣＬＩＰに限るものではなく、画像及びテキストの類似度を出力する構成の学習モデルであれば、どのような学習モデルが採用されてもよく、例えば大規模汎用画像モデルが採用され得る。なお大規模汎用画像モデルは、基盤モデル（様々なタスクに利活用できるように、大量のデータで学習させた高性能な事前訓練モデル）と呼ばれるもののうち、特に画像と言語で訓練したモデルである。大規模汎用画像モデルは、タスク固有の訓練データなしに、汎用的な画像認識が可能である。大規模汎用画像モデルには、例えばＳＬＩＰ（Self-supervision meets Language-Image Pre-training）、ＤｅＣＬＩＰ（Data efficient Contrastive Language-Image Pre-training）、ＦＩＬＩＰ（Fine-grained Interactive Language-Image Pre-Training）及びＣｏＣａ（Contrastive Captioner）等がある。

【００２１】

画像ＤＢ５は、サーバ装置１が一又は複数のカメラ３から取得した複数の画像を記憶して蓄積するデータベースである。画像ＤＢ５は、プログラム１２ａ及び学習モデル記憶部１２ｂ等が設けられる記憶部１２とは別の記憶部（記憶装置）に設けられていてもよい。画像ＤＢ５は、例えばカメラ３が撮影した画像と共に、この画像を撮影したカメラ３のＩＤ等の識別情報及びこの画像が撮影された日時等の情報を対応付けて記憶する。本実施の形態においては、カメラ３が撮影した画像に対してタイトル等のテキスト情報の付与、いわゆるタグ付けが行われることなく、画像ＤＢ５に画像が記憶されてよい。ただし、一部又は全部の画像について、タグ付けが行われて画像ＤＢ５に記憶されてもよい。

【００２２】

通信部１３は、例えばインターネット、ＬＡＮ（Local Area Network）又は携帯電話通信網等を含むネットワークＮを介して、種々の装置との間で通信を行う。本実施の形態において通信部１３は、ネットワークＮを介して、カメラ３及び端末装置７との間で通信を行う。通信部１３は、処理部１１から与えられたデータを他の装置へ送信すると共に、他の装置から受信したデータを処理部１１へ与える。

【００２３】
なお記憶部１２は、サーバ装置１に接続された外部記憶装置であってよい。またサーバ装置１は、複数のコンピュータを含んで構成されるマルチコンピュータであってよく、ソフトウェアによって仮想的に構築された仮想マシンであってもよい。またサーバ装置１は、上記の構成に限定されず、例えば可搬型の記憶媒体に記憶された情報を読み取る読取部、操作入力を受け付ける入力部、又は、画像を表示する表示部等を含んでもよい。
【００２４】
また本実施の形態に係るサーバ装置１では、記憶部１２に記憶されたプログラム１２ａを処理部１１が読み出して実行することにより、画像取得部１１ａ、テキスト取得部１１ｂ、類似度算出部１１ｃ、閾値決定部１１ｄ及び画像抽出部１１ｅ等が、ソフトウェア的な機能部として処理部１１に実現される。なお本図においては、処理部１１の機能部として、カメラ３が撮影した画像を扱う処理に関連する機能部を図示し、これ以外の処理に関する機能部は図示を省略している。
【００２５】
画像取得部１１ａは、通信部１３にてカメラ３との通信を行うことによって、カメラ３が撮影した画像を取得する処理を行う。例えば画像取得部１１ａは、カメラ３から画像が送信されるのを待機し、カメラ３から送信された画像をその都度受信することで、画像を受動的に取得してもよい。また例えば画像取得部１１ａは、１時間に１回又は１日に１回等の所定の周期でカメラ３に画像の送信を要求し、この要求に応じてカメラ３が送信する画像を受信することで、画像を能動的に取得してもよい。画像取得部１１ａは、カメラ３から取得した画像を、例えばカメラ３の識別情報及び撮影日時等の情報と共に画像ＤＢ５に記憶する。
【００２６】
テキスト取得部１１ｂは、通信部１３にて端末装置７との通信を行うことによって、ユーザが端末装置７に入力したキーワード等のテキストの情報を取得する処理を行う。テキスト取得部１１ｂは、例えば端末装置７から画像の送信要求と共に与えられるテキストを通信部１３にて受信することによって、テキストを取得する。テキスト取得部１１ｂは、取得したテキストを記憶部１２に一時的に記憶する。
【００２７】
類似度算出部１１ｃは、画像ＤＢ５に記憶された各画像とテキスト取得部１１ｂが取得したテキストとの類似度を算出する処理を行う。本実施の形態に置いて類似度算出部１１ｃは、学習モデル記憶部１２ｂに記憶された学習済の学習モデルを用いて、画像及びテキストの類似度を算出する。本実施形態に係る学習モデルは、画像及びテキストの入力を受け付けて、この画像及びテキストの類似度を出力するように予め機械学習がなされた学習モデルである。類似度算出部１１ｃは、画像ＤＢ５に記憶された画像とテキスト取得部１１ｂが取得したテキストとを学習モデルへ入力し、学習モデルが出力する類似度を取得することで、画像及びテキストの類似度を算出する。
【００２８】
本実施の形態において類似度算出部１１ｃは、画像ＤＢ５に記憶された複数の画像のうち、処理対象となる画像の全てについて、画像及びテキストの類似度を算出する。例えばユーザが撮影日時又は撮影場所等の条件を設定した場合には、画像ＤＢ５に記憶された全画像のうち設定された条件に合致する画像が、処理対象の画像となる。例えばユーザがこれらの条件を設定しない場合、画像ＤＢ５に記憶された全ての画像が処理対象の画像となり得る。類似度算出部１１ｃは、例えば処理対象の画像がＮ個である場合、１つのテキストと各画像との類似度として、Ｎ個の類似度を算出する。
【００２９】
閾値決定部１１ｄは、類似度算出部１１ｃが算出した類似度と比較する閾値、即ち画像及びテキストが類似しているか否かを判定するための閾値を決定する処理を行う。例えば、画像及びテキストの類似度が閾値を超える場合に、この画像及びテキストが類似していると判定される。閾値決定部１１ｄによる閾値の決定方法の詳細は、後述する。

【００３０】
画像抽出部１１ｅは、画像ＤＢ５に記憶された複数の画像の中から、テキスト取得部１１ｂが取得したテキストに類似する画像を抽出する処理を行う。画像抽出部１１ｅは、類似度算出部１１ｃが算出した類似度及び閾値決定部１１ｄが決定した閾値を比較し、類似度が閾値を超える画像及びテキストの組を特定する。画像抽出部１１ｅは、特定した組の画像を画像ＤＢ５から読み出すことにより、テキストに類似する画像を抽出する。画像抽出部１１ｅは、抽出した一又は複数の画像をテキストの送信元の端末装置７へ送信し、送信した一又は複数の画像を検索結果又は抽出結果として端末装置７の表示部に表示させる。
【００３１】
＜学習モデル＞
図３は、本実施の形態に係る情報処理システムが使用する学習モデル２０の一構成例を示す模式図である。本実施の形態に係る学習モデル２０は、画像及びテキストの入力を受け付けて、この画像及びテキストの類似度を出力する学習モデルである。学習モデル２０には、例えばＣＬＩＰの学習モデルが採用され得る。学習モデル２０は、テキストエンコーダ２１及び画像エンコーダ２２を有しており、入力されたテキストをテキストエンコーダ２１へ入力し、入力された画像を画像エンコーダ２２へ入力する。
【００３２】
テキストエンコーダ２１は、入力されたテキストを所定次元の特徴量のベクトルに変換して出力する。同様に、画像エンコーダ２２は、入力された画像を所定次元の特徴量のベクトルに変換して出力する。テキストエンコーダ２１は、例えばTransformer又はＲＮＮ（Recurrent Neural Network）等の構成が採用され得る。画像エンコーダ２２は、例えばVision Transformer又はＣＮＮ（Convolutional Neural Network）等の構成が採用され得る。学習モデル２０は、テキストエンコーダ２１が出力する特徴量のベクトルと、画像エンコーダ２２が出力する特徴量のベクトルとに基づいて、類似度を算出して出力する。例えば学習モデル２０は、２つの特徴量のベクトルの内積を算出し、算出した内積の値を類似度として出力する。
【００３３】
図４は、学習モデル２０の学習方法の概要を説明するための模式図である。学習モデル２０の機械学習を行うために、画像に対してテキストが対応付けられた複数の学習用のデータの収集がなされる。このデータは、例えば犬の画像に対して「犬」のテキストを対応付けた一組のデータである。図４に示す例では、Ｎ個（Ｎ組）の学習用データが機械学習に用いられており、画像１及びテキスト１が対応する組であり、画像２及びテキスト２が対応する組であり、…、画像Ｎ及びテキストＮが対応する組である。これらＮ組の学習用データに含まれるＮ個の画像を画像エンコーダ２２へ入力することで、Ｎ個の画像に対するＮ個の特徴量が得られる。図４においては画像１の特徴量をＧ１、画像２の特徴量をＧ２、…、画像Ｎの特徴量をＧＮと記載している。同様にして、Ｎ組の学習用データに含まれるＮ個のテキストをテキストエンコーダ２１へ入力することで、Ｎ個のテキストに対するＮ個の特徴量が得られる。図４においてはテキスト１の特徴量をＴ１、テキスト２の特徴量をＴ２、…、テキストＮの特徴量をＴＮと記載している。
【００３４】
画像エンコーダ２２が出力する特徴量のベクトルと、テキストエンコーダ２１が出力する特徴量のベクトルとの内積を算出することで、画像及びテキストの類似度が算出できる。Ｎ個の画像１～Ｎを基に得られるＮ個の特徴量Ｇ１～ＧＮと、Ｎ個のテキスト１～Ｎを基に得られるＮ個の特徴量Ｔ１～ＴＮとの組み合わせから、Ｎ×Ｎ個の類似度を算出することができる。図４においては、画像１及びテキスト１の類似度をＧ１・Ｔ１、画像１及びテキスト２の類似度をＧ１・Ｔ２、…、画像Ｎ及びテキストＮの類似度をＧＮ・ＴＮと記載している。例えば画像１及びテキスト１の組み合わせは、本来の正しい組み合わせ（正例）であり、類似度が高いことが期待される。これに対して、画像１及びテキスト２の組み合わせは、本来とは異なる誤った組み合わせ（負例）であり、類似度が低いことが期待される。

【００３５】
そこで、ｉ番目の画像ｉとｊ番目のテキストｊとの類似度をＧｉ・Ｔｊとし、ｉ＝ｊの場合の特徴量に対する正解ラベル（教師ラベル、正解値等）を「１」とし、ｉ≠ｊの場合の特徴量に対する正解ラベルを「０」として機械学習を学習モデル２０に対して行うことにより、テキストエンコーダ２１及び画像エンコーダ２２の内部のパラメータを決定することができる。機械学習は、例えば勾配降下法、確率的勾配降下法又は誤差逆伝播法等の手法を用いて行われ得る。機械学習は、既存の技術であるため、詳細な説明は省略する。
【００３６】
即ち、上述の学習モデル２０の機械学習では、Ｎ組の画像及びテキストの正例のデータから、Ｎ×（Ｎ－１）組の負例のデータを生成し、正例のデータから算出される類似度の正解ラベルに「１」を与え、負例のデータから算出される類似度に正解ラベル「０」を与えて、Ｎ×Ｎ個の正解ラベルを用いた機械学習が行われる。
【００３７】
なお、学習モデル２０を生成するための上述の機械学習の処理は、サーバ装置１が行うのではなく、別の装置にて行われてよい。サーバ装置１は、機械学習がなされた学習済の学習モデル２０を別の装置から取得して学習モデル記憶部１２ｂに記憶する。サーバ装置１は、例えば端末装置７からテキストの入力を伴う画像の検索又は抽出等の要求が与えられた場合に、学習モデル記憶部１２ｂに記憶した学習済の学習モデル２０を用いて、画像ＤＢ５に記憶された画像の中からテキストに類似する画像を抽出して端末装置７へ送信する。
【００３８】
図５は、本実施の形態に係るサーバ装置１が行う画像抽出処理の手順を示すフローチャートである。本実施の形態に係るサーバ装置１の処理部１１は、端末装置７から画像抽出を行う要求を受信したか否かを判定する（ステップＳ１）。画像抽出を行う要求を受信していない場合（Ｓ１：ＮＯ）、処理部１１は、要求を受信するまで待機する。要求を受信した場合（Ｓ１：ＹＥＳ）、処理部１１のテキスト取得部１１ｂは、要求と共に端末装置７から送信される抽出条件となるテキストを取得する（ステップＳ２）。
【００３９】
処理部１１の画像取得部１１ａは、画像ＤＢ５に記憶された処理対象の複数の画像から１つの画像を取得する（ステップＳ３）。処理部１１の類似度算出部１１ｃは、ステップＳ３にて取得した画像及びステップＳ２にて取得したテキストを、学習モデル記憶部１２ｂに記憶された学習済の学習モデル２０へ入力する（ステップＳ４）。類似度算出部１１ｃは、画像及びテキストの入力に応じて学習モデル２０が出力する類似度を取得する（ステップＳ５）。
【００４０】
なお本フローチャートにおいては、１つの画像及び１つのテキストを学習モデルへ入力して１つの類似度を取得しているが、これに限るものではない。いわゆるバッチ処理により、例えば複数の画像及び１つのテキストを学習モデルへ入力し、各画像とテキストとの複数の類似度を取得してもよい。このようなバッチ処理を採用することによって、画像抽出処理の高速化が期待できる。
【００４１】
画像抽出部１１ｅは、ステップＳ５にて取得した画像及びテキストの類似度が、予め定められた閾値を超えるか否かを判定する（ステップＳ６）。類似度が閾値を超える場合（Ｓ６：ＹＥＳ）、画像抽出部１１ｅは、この類似度に対応する画像を要求元の端末装置７へ送信し（ステップＳ７）、ステップＳ８へ処理を進める。類似度が閾値を超えない場合（Ｓ６：ＮＯ）、画像抽出部１１ｅは、画像を送信せずに、ステップＳ８へ処理を進める。
【００４２】
処理部１１は、画像ＤＢ５に記憶された画像のうち、処理対象とする複数の画像の全てについてステップＳ３～Ｓ７の処理を終了したか否かを判定する（ステップＳ８）。全ての画像について処理を終了していない場合（Ｓ８：ＮＯ）、処理部１１は、ステップＳ３

へ処理を戻し、別の画像を取得して同様の処理を繰り返し行う。全ての画像について処理を終了した場合（Ｓ８：ＹＥＳ）、処理部１１は、画像抽出の処理を終了する。

【００４３】

＜閾値の決定方法＞

上述のように、本実施の形態に係る情報処理システムは、学習モデル２０が出力する画像及びテキストの類似度が閾値を超える場合に、この画像がテキストに類似する画像であると判定する。この判定に用いられる閾値の決定方法には、例えば以下の４つの方法のいずれかが採用され得る。

（１）適合度又は再現度に基づく閾値の決定

（２）代表値に基づく閾値の決定

（３）分布に基づく閾値の決定

（４）ユーザによる閾値の決定

【００４４】

（１）適合度又は再現度に基づく閾値の決定

第１の決定方法には、画像に対して正しいテキストが対応付けられたデータ（正例）と、画像に対して誤ったテキストが対応付けられたデータ（負例）とを含む、検証用データが必要である。本実施の形態に係るサーバ装置１は、機械学習がなされた学習モデル２０に対して検証用データを入力し、検証用データの画像及びテキストの各組に対する類似度を取得する。サーバ装置１は、閾値Ｘを用いて類似度との比較を行った場合の適合度又は再現度を、閾値Ｘの値を変化させてそれぞれ算出し、適合度又は再現度が所定値（例えば０．９）となる閾値Ｘを特定する。なお、適合度又は再現度に対する所定値は、本実施の形態に係る情報処理システムの設計者又は管理者等により予め定められる。

【００４５】

なお適合度は、類似度及び閾値Ｘの比較の結果から正例と予想されたデータのうち、実際に正例だったデータの割合である。また再現度は、正例の真値のうち、正しく予想されたデータの割合である。機械学習モデルの適合度又は再現度の算出方法は、既存の技術であるため、詳細な説明を省略する。

【００４６】

図６は、適合度に基づく閾値の決定方法を説明するための模式図である。図６の上段に記載のグラフは、検証用データの正例及び負例について類似度の分布を示すヒストグラムであり、横軸を類似度とし、縦軸をデータ数としている。図６の下段に記載のグラフは、各類似度を閾値とした場合の検証用データの適合度を示すグラフであり、横軸を類似度（閾値）とし、縦軸を適合度としている。下段のグラフに描かれた破線の水平線は、適合度＝０．９を示しており、例えば設計者又は管理者等が適合度＝０．９を閾値の条件として決定したことを示している。サーバ装置１は、適合度が０．９となる類似度を閾値として決定し、決定した閾値を例えば学習モデル記憶部１２ｂに学習モデル２０に関する情報と共に記憶し、図５に示した画像抽出処理において記憶した閾値を用いて判定を行う。

【００４７】

図７は、再現度に基づく閾値の決定方法を説明するための模式図である。図７の上段に記載のグラフは、図６の上段に記載したグラフと同じものであり、検証用データの正例及び負例について類似度の分布を示すヒストグラムである。図７の下段に記載のグラフは、各類似度を閾値とした場合の検証用データの再現度を示すグラフであり、横軸を類似度（閾値）とし、縦軸を再現度としている。下段のグラフに描かれた破線の水平線は、再現度＝０．９を示しており、例えば設計者又は管理者等が再現度＝０．９を閾値の条件として決定したことを示している。サーバ装置１は、再現度が０．９となる類似度を閾値として決定し、決定した閾値を例えば学習モデル記憶部１２ｂに学習モデル２０に関する情報と共に記憶し、図５に示した画像抽出処理において記憶した閾値を用いて判定を行う。

【００４８】

なおサーバ装置１は、適合度に基づく閾値の決定又は再現度に基づく閾値の決定の少なくとも一方を行って閾値を決定すればよい。いずれの方法で閾値を決定するかは、例えば

設計者又は管理者等により予め定められ得る。又は、両方法でそれぞれ閾値を決定しておき、ユーザがいずれの閾値を採用するかを端末装置７にて選択することが可能であってもよい。

【００４９】

（２）代表値に基づく閾値の決定

第２の決定方法には、画像に対して正しいテキストが対応付けられたデータ（正例）が検証用データとして用いられる。この検証用データには、画像に対して誤ったテキストが対応付けられたデータ（負例）が含まれない。本実施の形態に係るサーバ装置１は、機械学習がなされた学習モデル２０に対して検証用データを入力し、検証用データの画像及びテキストの各組に対する類似度を取得する。サーバ装置１は、正例の検証用データに関して算出した複数の類似度について、例えば平均値又は最小値等の代表値を算出し、算出した代表値を閾値とする。サーバ装置１は、算出した代表値を閾値として例えば学習モデル記憶部１２ｂに学習モデル２０に関する情報と共に記憶し、図５に示した画像抽出処理において記憶した閾値を用いて判定を行う。

【００５０】

図８は、代表値に基づく閾値の決定方法を説明するための模式図である。図８の上段に記載のグラフは、正例の検証用データについて類似度の分布を示すグラフであり、横軸を類似度とし、縦軸をデータ数としている。図８の下段に記載のグラフは、画像抽出の対象となる全画像について類似度の分布を示すグラフであり、横軸を類似度とし、縦軸をデータ数としている。図８において破線で示す垂直線は検証用データの類似度の平均値を示し、一点鎖線で示す垂直線は検証用データの類似度の最小値を示している。サーバ装置１は、検証用データの類似度の平均値又は最小値等の代表値を算出して閾値として用いることにより、この閾値より類似度が大きい画像が抽出される。

【００５１】

なお代表値を閾値として平均値又は最小値等のいずれを採用するかは、例えば本実施の形態に係る情報処理システムの設計者又は管理者等により予め定められる。また代表値は、類似度の平均値又は最小値に限らず、これら以外の値が採用されてもよい。またサーバ装置１は、例えば負例の検証用データを用いて類似度を取得し、取得した複数の類似度の平均値又は最大値等の代表値を閾値として用いてもよい。

【００５２】

（３）分布に基づく閾値の決定

第３の決定方法では、正例又は負例の検証用データを用いるのではなく、画像抽出の対象となる全画像に対して指定されたテキストとの類似度をそれぞれ取得し、取得した全類似度の分布に基づいて閾値を決定する。サーバ装置１は、例えば画像抽出の対象となる全画像に対して、例えば設計者又は管理者等により類似度の上位Ｘ％を抽出することが定められている。サーバ装置１は、全画像について取得した画像及びテキストの類似度をソートして並べ替え、類似度が高いものから上位Ｘ％に相当する類似度を特定し、特定した類似度を閾値とする。

【００５３】

図９は、分布に基づく閾値の決定を説明するための模式図である。図９に記載のグラフは、画像抽出の対象となる全画像について類似度の分布を示すグラフであり、横軸を類似度とし、縦軸をデータ数としている。図８において破線で示す垂直線は、例えば類似度が高いものから上位５％に相当する類似度を示しており、個の類似度が閾値として採用される。なお、上位５％は一例であって、これに限るものではない。

【００５４】

ただしサーバ装置１は、類似度のソートを行うのではなく、類似度の分布が正規分布に従うものと仮定して近似的に閾値を決定してもよい。サーバ装置１は、対象の全画像について算出した複数の類似度について平均、分散及び標準偏差を算出する。サーバ装置１は、例えば正規分布における累積確率（１００％－Ｘ％）／１００と、算出した平均及び標準偏差とを基に、累積正規分布の逆関数を用いてＸ％に相当する類似度を算出し、この類

似度を閾値とすることができる。

【００５５】

なお、ソートにより上位Ｘ％の類似度を特定する方法と、類似度が正規分布に従うと仮定した近似による類似度を算出する方法とのいずれを採用するかは、例えば本実施の形態に係る情報処理システムの設計者又は管理者等により予め定められる。数値のソート、及び、正規分布に基づく近似値の算出等は、既存の技術であるため、詳細な手順の説明を省略する。またサーバ装置１は、正規分布以外の分布、例えばベータ分布等の他の分布に近似して近似値を算出してもよい。

【００５６】

（４）ユーザによる閾値の決定

第４の決定方法では、ユーザが端末装置７にて閾値を決定することができる。サーバ装置１は、端末装置７を介してユーザが入力したテキストを取得し、対象となる全ての画像と取得したテキストとの類似度をそれぞれ学習モデル２０を用いて取得する。サーバ装置１は、全ての画像及びテキストの組について算出した類似度について例えばヒストグラム等のグラフを作成し、作成したグラフのデータを端末装置７へ送信する。またサーバ装置１は、ヒストグラムのデータと共に、デフォルトの閾値を用いて抽出した画像を端末装置７へ送信する。サーバ装置１からグラフのデータ及び抽出された画像を受信した端末装置７は、受信したデータに基づいて、ヒストグラム等のグラフを表示部に表示すると共に、抽出された一又は複数の画像を表示する。

【００５７】

図１０は、ユーザによる閾値の決定方法を説明するための模式図である。本実施の形態に係る端末装置７は、ユーザからテキストの入力を受け付けてサーバ装置１へ送信した後、サーバ装置１から送信されるデータを受信してヒストグラム等のグラフを表示部に表示する。図１０に示す例では、端末装置７は、画面の左上の領域にグラフを表示している。このグラフは、横軸を類似度とし、縦軸をデータ数としたヒストグラムである。また端末装置７は、サーバ装置１から抽出結果として送信される一又は複数の画像を受信して表示部に表示する。図１０に示す例では、端末装置７は、画面の右側の領域に、複数の画像をマトリクス状に並べて表示している。複数の画像は、例えば類似度の大きい／小さい順、又は、撮影日時が新しい／古い順等の適宜の順番で並べて表示される。

【００５８】

端末装置７は、このヒストグラムに対して、破線で示す垂直線を、閾値を示す指標として重ねて表示する。閾値の指標はまず予め定められたデフォルト値で表示され、ユーザは例えばマウス又はタッチパネル等の入力装置を利用してこの指標を水平方向に移動させることによって、閾値の設定を増減することができる。ユーザの操作により閾値が変更された場合、端末装置７は、変更後の閾値をサーバ装置１へ送信する。サーバ装置１は、端末装置７から変更後の閾値を受信し、この閾値を用いて画像の再抽出を行い、抽出結果を端末装置７へ送信する。端末装置７は、変更された閾値に基づく抽出結果をサーバ装置１から受信し、表示部に並べて表示する画像を新たに受信した画像に更新する。

【００５９】

なお、デフォルトの閾値は、例えば本実施の形態に係る情報処理システムの設計者又は管理者等により予め定められ得る。また例えば端末装置７は、前回にユーザが設定した閾値を記憶しておき、記憶した閾値をデフォルトの閾値としてサーバ装置１へ送信してもよい。

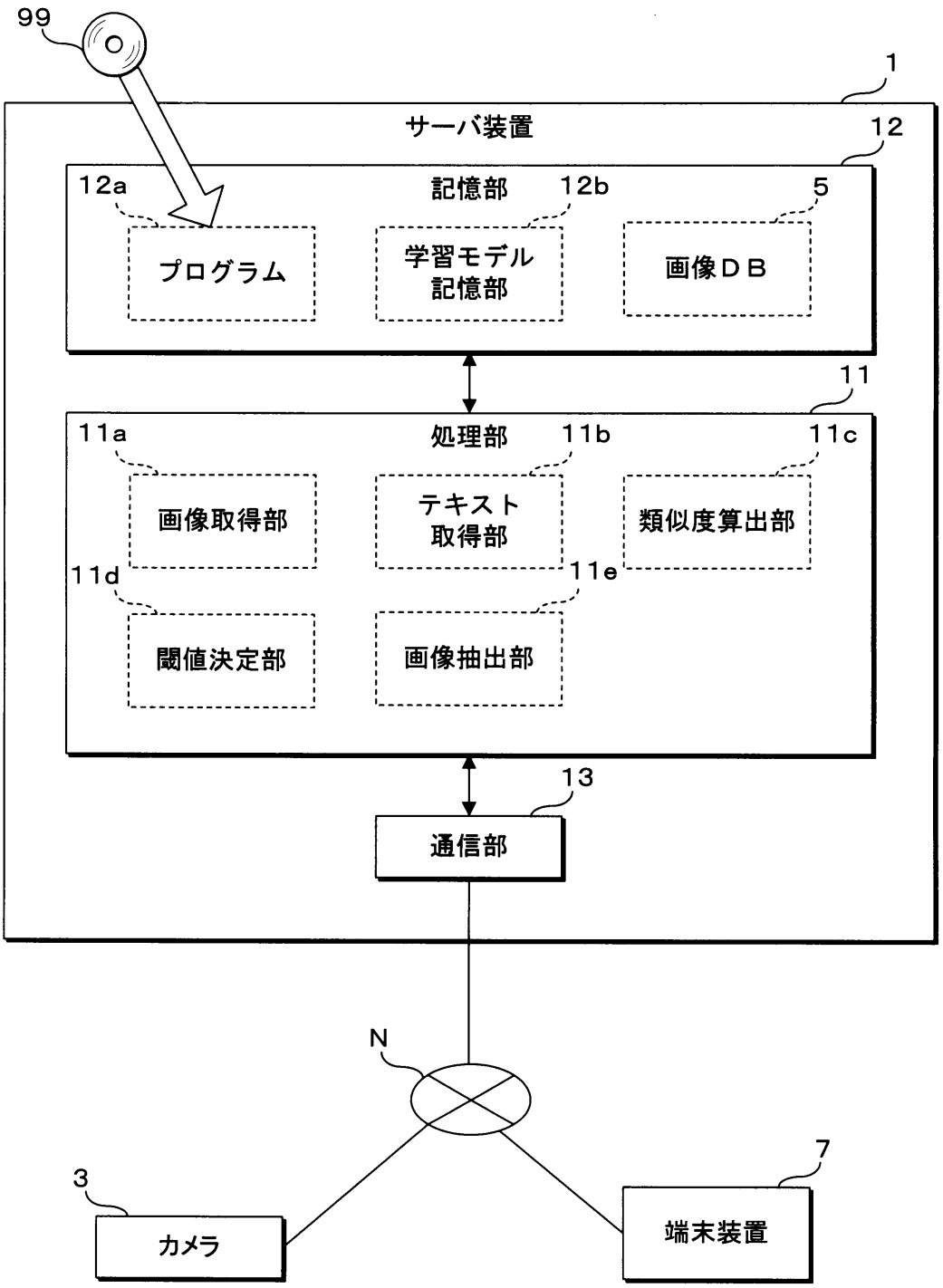
【００６０】

またユーザによる閾値の決定を受け付ける方法は、上記のヒストグラム等のグラフを用いる方法に限らず、種々の方法が採用され得る。例えば、類似度の最小値から最大値までの間で数値設定を受け付けるスライダー又はバー等を表示して、端末装置７がこれらのスライダー又はバー等に対するユーザの操作を受け付けて閾値を決定してもよい。また例えば、ユーザが閾値とする数値を直接的に入力し、端末装置７が入力された数値を取得して閾値としてもよい。

【００６１】
＜まとめ＞
以上の構成の本実施の形態に係る情報処理システムでは、サーバ装置１が画像ＤＢ５から処理対象となる複数の画像を取得し、画像の抽出条件となるテキストを端末装置７から取得し、予め機械学習がなされた学習モデル２０に書く画像及びテキストを入力して類似度を取得することにより、複数の画像とテキストとの各組の類似度を取得する。サーバ装置１は、画像及びテキストの各組の類似度と所定の閾値とを比較して、処理対象の複数の画像から類似度が閾値を超える画像を抽出して出力する。これにより本実施の形態に係る情報処理システムでは、画像ＤＢ５に記憶する画像に予めタグ付けを行う必要なく、画像ＤＢ５に記憶した複数の画像からテキスト入力に基づく画像の抽出又は検索等を行うことが期待できる。
【００６２】
また本実施の形態に係る情報処理システムでは、サーバ装置１が、画像及びテキストの正例の組及び負例の組をそれぞれ複数取得し、各組について学習モデル２０による類似度を取得し、取得した類似度に基づいて適合度又は再現度を算出し、算出した適合度又は再現度に基づいて閾値を決定する。これにより本実施の形態に係る情報処理システムでは、例えば正例及び負例の検証用データが利用できる場合に、予め機械学習がなされた学習モデル２０の性能又は特性等に適した閾値を決定することが期待できる。
【００６３】
また本実施の形態に係る情報処理システムでは、サーバ装置１が、画像及びテキストの正例の組を複数取得し、正例の各組について学習モデル２０による類似度を取得し、取得した類似度の分布に基づいて閾値を決定する。サーバ装置１は、例えば類似度の分布に関する平均値又は最小値等の代表値を算出し、算出した代表値を閾値とすることができる。これにより本実施の形態に係る情報処理システムでは、例えば正例の検証用データが利用できる場合に、学習モデル２０の正例に対する類似度の算出の特性に適した閾値を決定することが期待できる。
【００６４】
また本実施の形態に係る情報処理システムでは、サーバ装置１が、正解値のラベル又はタグ等が付与されていない画像を用いて、テキストとの類似度を学習モデル２０にて取得し、複数の画像についての類似度の分布に基づいて閾値を決定する。これにより本実施の形態に係る情報処理システムは、画像抽出の対象となる画像ＤＢ５に記憶された複数の画像を基に閾値を決定することができるため、実際に画像ＤＢ５に記憶された画像の特性等に適した閾値を決定することが期待できる。
【００６５】
また本実施の形態に係る情報処理システムでは、サーバ装置１が、複数の画像について算出した類似度に基づいて画像を順位付け（ソート）し、例えば上位Ｘ％等の所定の順位の画像を抽出するように閾値を決定する。これにより本実施の形態に係る情報処理システムでは、類似度が高いものを優先して必要な量だけ抽出することができる。
【００６６】
また本実施の形態に係る情報処理システムでは、サーバ装置１が、類似度の分布に関する例えば平均、分散又は標準偏差等のパラメータを算出し、類似度の分布が例えば正規分布などの所定分布であるとみなして、パラメータに応じた閾値を決定する。これにより本実施の形態に係る情報処理システムでは、画像ＤＢ５に記憶された複数の画像について算出した類似度を基に、演算量が少ない方法で閾値を決定することが期待できる。
【００６７】
また本実施の形態に係る情報処理システムでは、サーバ装置１が、画像及びテキストの各組について学習モデル２０が出力した類似度の分布を例えば端末装置７の表示部に表示させ、端末装置７を介してユーザからの閾値の設定を受け付ける。これにより本実施の形態に係る情報処理システムは、ユーザの好み等に適した閾値を用いて画像の抽出を行うことが期待できる。

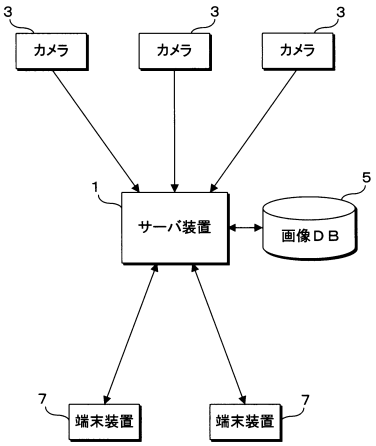
【００６８】
　また本実施の形態に係る情報処理システムでは、学習モデル２０は、入力された画像の特徴量を出力する画像エンコーダ２２と、入力されたテキストの特徴量を出力するテキストエンコーダ２１と、画像エンコーダ２２が出力した特徴量及びテキストエンコーダ２１が出力した特徴量を基に類似度を算出する算出部とを備える構成である。学習モデル２０には、例えば大規模汎用画像モデルであるＣＬＩＰの学習モデルが採用され得る。これにより本実施の形態に係る情報処理システムでは、画像及びテキストの類似度を精度よく算出することが期待できる。
【００６９】
　今回開示された実施形態はすべての点で例示であって、制限的なものではないと考えられるべきである。本発明の範囲は、上記した意味ではなく、特許請求の範囲によって示され、特許請求の範囲と均等の意味及び範囲内でのすべての変更が含まれることが意図される。
【００７０】
　各実施形態に記載した事項は相互に組み合わせることが可能である。また、特許請求の範囲に記載した独立請求項及び従属請求項は、引用形式に関わらず全てのあらゆる組み合わせにおいて、相互に組み合わせることが可能である。さらに、特許請求の範囲には他の２以上のクレームを引用するクレームを記載する形式（マルチクレーム形式）を用いているが、これに限るものではない。マルチクレームを少なくとも１つ引用するマルチクレーム（マルチマルチクレーム）を記載する形式を用いて記載してもよい。
【符号の説明】
【００７１】
１　サーバ装置（情報処理装置、コンピュータ）
３　カメラ
５　画像ＤＢ
７　端末装置
１１　処理部
１１ａ　画像取得部
１１ｂ　テキスト取得部
１１ｃ　類似度算出部
１１ｄ　閾値決定部
１１ｅ　画像抽出部
１２　記憶部
１２ａ　プログラム（コンピュータプログラム）
１２ｂ　学習モデル記憶部
１３　通信部
２０　学習モデル
２１　テキストエンコーダ
２２　画像エンコーダ
９９　記録媒体
Ｎ　ネットワーク

【要約】
【課題】テキストに基づく画像の検索及び抽出等を実現することが期待できる情報処理方法、コンピュータプログラム及び情報処理装置を提供する。
【解決手段】本実施の形態に係る情報処理方法は、情報処理装置が、処理対象となる複数の画像を取得し、前記複数の画像からの画像の抽出条件となるテキストを取得し、画像及びテキストの入力を受け付けて前記画像及び前記テキストの類似度を出力するよう機械学習がなされた学習モデルへ、取得した画像及びテキストを入力して前記学習モデルが出力する前記画像及び前記テキストの類似度を取得することで、前記複数の画像と前記テキストとの各組の類似度を取得し、取得した各組の類似度と所定の閾値とを比較し、前記複数の画像から、前記類似度が前記閾値を超える画像を抽出して出力する。
【選択図】図２

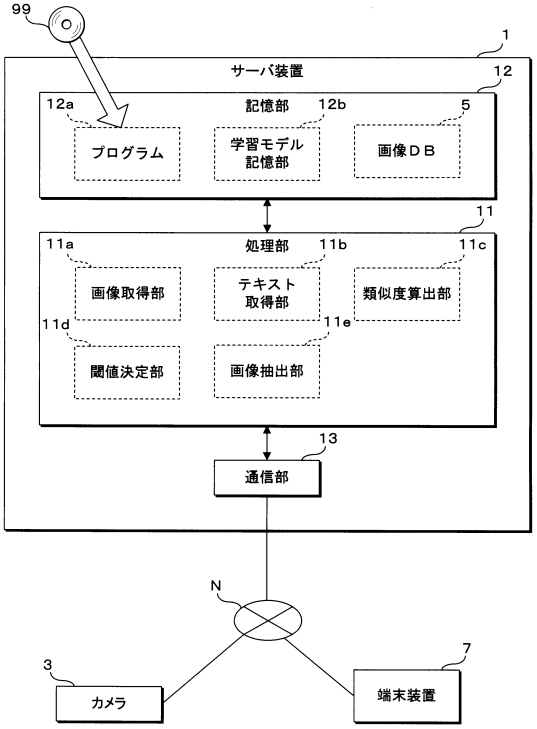

99
1
サーバ装置
12
12a
記憶部
12b
5
プログラム
学習モデル
記憶部
画像ＤＢ
11
11a
処理部
11b
11c
画像取得部
テキスト
取得部
類似度算出部
11d
11e
閾値決定部
画像抽出部
13
通信部
N
3
7
カメラ
端末装置

【図面】

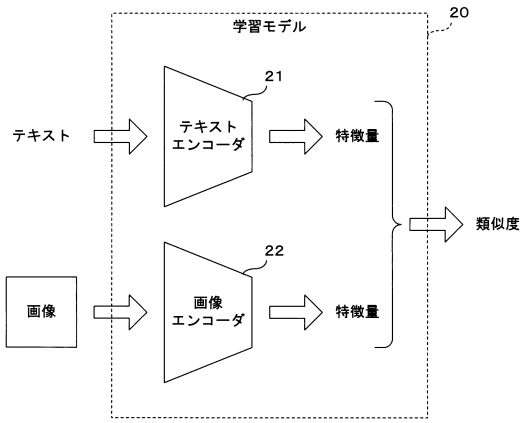
【図１】


3
カメラ
3
カメラ
3
カメラ
1
サーバ装置
5
画像ＤＢ
7
端末装置
7
端末装置

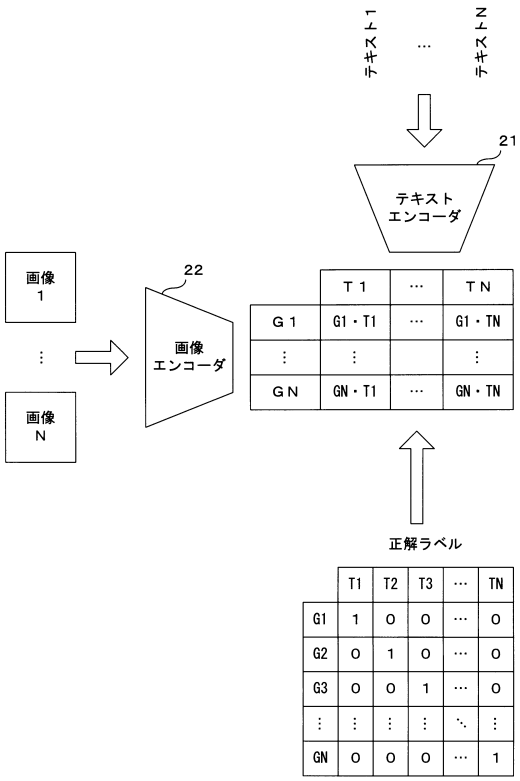
【図２】


99
1
サーバ装置
12
12a
記憶部
12b
5
プログラム
学習モデル
記憶部
画像ＤＢ
11
11a
処理部
11b
11c
画像取得部
テキスト
取得部
類似度算出部
11d
11e
閾値決定部
画像抽出部
13
通信部
N
3
カメラ
7
端末装置

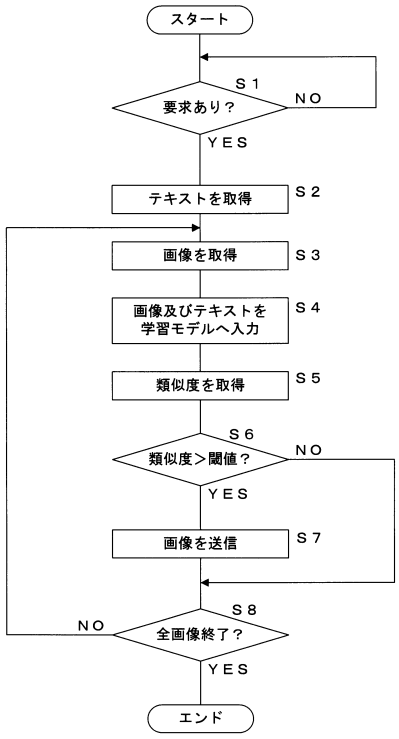
【図３】


学習モデル
20
21
テキスト
テキスト
エンコーダ
特徴量
類似度
22
画像
画像
エンコーダ
特徴量

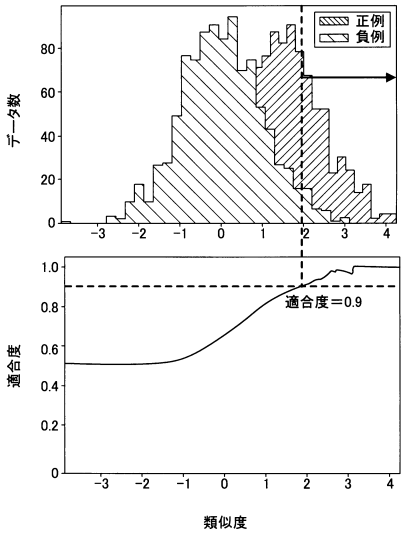
【図４】


テキスト1
…
テキストN
21
テキスト
エンコーダ
画像
1
画像
N
22
画像
エンコーダ
T１
…
TＮ
Ｇ１
G1・T1
…
G1・TN
ＧＮ
GN・T1
…
GN・TN
正解ラベル
T1
T2
T3
…
TN
G1
1
0
0
…
0
G2
0
1
0
…
0
G3
0
0
1
…
0
GN
0
0
0
…
1

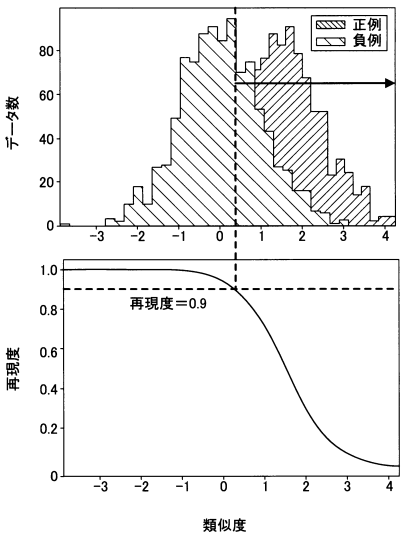
【図５】


スタート
S１
要求あり？
ＮＯ
ＹＥＳ
テキストを取得
S２
画像を取得
S３
画像及びテキストを
学習モデルへ入力
S４
類似度を取得
S５
S６
類似度＞閾値？
ＮＯ
ＹＥＳ
画像を送信
S７
S８
ＮＯ
全画像終了？
ＹＥＳ
エンド

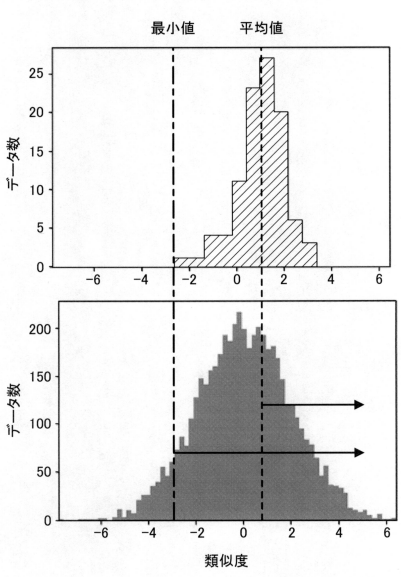
【図６】


正例
負例
データ数
適合度＝0.9
適合度
類似度

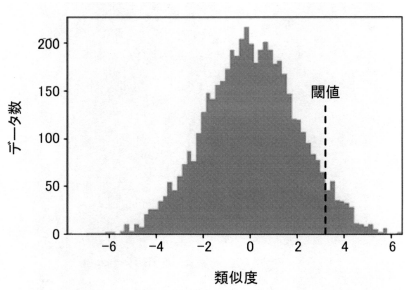
【図７】


正例
負例
データ数
再現度＝0.9
再現度
類似度

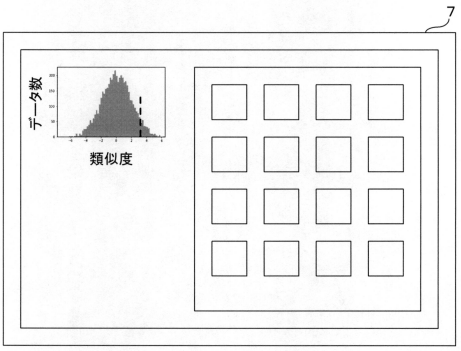
【図８】


最小値
平均値
データ数
データ数
類似度

【図９】



【図１０】



フロントページの続き

会社エクサウィザーズ内

審査官　甲斐 哲雄

(56)参考文献　特表２０２０－５２２７９１（ＪＰ，Ａ）
特表２０２２－５０９３２７（ＪＰ，Ａ）
特開２０２２－１８０９４１（ＪＰ，Ａ）

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１６／００－１６／９５８