

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7575813号
(P7575813)

(45)発行日 令和6年10月30日(2024.10.30)

(24)登録日 令和6年10月22日(2024.10.22)

(51)国際特許分類

F I

G 0 6 N 20/00 (2019.01)

G 0 6 N 20/00 1 3 0

請求項の数 8 (全17頁)

(21)出願番号	特願2023-19540(P2023-19540)	(73)特許権者	520185487
(22)出願日	令和5年2月10日(2023.2.10)		株式会社アートクリフ
(65)公開番号	特開2024-114102(P2024-114102 A)		東京都千代田区神田佐久間河岸7 8 - 6 第二寿ビルディング4階
(43)公開日	令和6年8月23日(2024.8.23)	(74)代理人	100122426
審査請求日	令和5年2月13日(2023.2.13)		弁理士 加藤 清志
		(72)発明者	鎌田 佑
			東京都千代田区神田佐久間河岸7 8 - 6 第二寿ビルディング4 F 株式会社ア ートクリフ内
		(72)発明者	松尾 光浩
			東京都千代田区神田佐久間河岸7 8 - 6 第二寿ビルディング4 F 株式会社ア ートクリフ内
		審査官	石坂 知樹

最終頁に続く

(54)【発明の名称】 文書生成システム、文書生成方法およびプログラム

(57)【特許請求の範囲】

【請求項1】

文書データとしての学習データを収集する学習データ収集部と、
該収集された前記学習データの中から特定の前記学習データを教師データとして抽出する抽出部と、
該抽出された前記教師データを入力し、第1の機械学習を実行して、第1の学習済モデルを生成する第1の学習済モデル生成部と、
該生成された前記第1の学習済モデルを用いて、前記学習データから選別データを生成する選別データ生成部と、
該生成された前記選別データとキーワードとを入力し、第2の機械学習を実行して、第2の学習済モデルを生成する第2の学習済モデル生成部と、
前記第2の学習済モデルとキーワードとを用いて、推論結果としての文書データを出力する推論結果出力部と、
前記第1の学習済モデルを用いて、推論結果としての文書データを評価し、評価の高い文書データを出力する文書データ出力部と、
を備え、
前記教師データは、前記特定の前記学習データにラベルが付されたデータであり、前記ラベルには、文書に種別に応じた複数のパラメータに対する充足度が示されており、
前記第1の学習済モデル生成部は、前記教師データとして、前記特定の前記学習データとラベルとを入力し、

10

前記選別データ生成部は、前記第1の学習済モデルを用いて、前記特定の学習データ以外の前記学習データに前記ラベルを付し、前記パラメータごとの充足度が高い前記ラベルが多く付された前記学習データを前記選別データとして生成することを特徴とする文書生成システム。

【請求項2】

前記第1の学習済モデル生成部および前記選別データ生成部は、前記特定の前記学習データと前記複数のパラメータのうち、1つの前記パラメータに対する充足度とを組み合わせた文書データに対して、処理を実行することを特徴とする請求項1に記載の文書生成システム。

【請求項3】

前記文書データ出力部から出力される前記文書データは、前記選別データ生成部の前記学習データとしてフィードバックされることを特徴とする請求項1に記載の文書生成システム。

【請求項4】

前記パラメータに対する充足度は、2値で示されることを特徴とする請求項1に記載の文書生成システム。

【請求項5】

前記文書データが求人票の募集要項であることを特徴とする請求項1から4のいずれか1項に記載の文書生成システム。

【請求項6】

前記パラメータが、自社の現状、やってほしい仕事、ポジション、仕事の進め方、将来像を含み、前記充足度は、各パラメータに関する記載が前記文書データに記載されているか否かを2値で示していることを特徴とする請求項4に記載の文書生成システム。

【請求項7】

学習データ収集部と、抽出部と、第1の学習済モデル生成部と、選別データ生成部と、第2の学習済モデル生成部と、推論結果出力部と、文書データ出力部と、を含む文書生成システムにおける文書生成方法であって、

前記学習データ収集部が、文書データとしての学習データを収集する第1の工程と、前記抽出部が、該収集された前記学習データの中から特定の前記学習データを教師データとして抽出する第2の工程と、

前記第1の学習済モデル生成部が、該抽出された前記教師データを入力し、第1の機械学習を実行して、第1の学習済モデルを生成する第3の工程と、

前記選別データ生成部が、該生成された前記第1の学習済モデルを用いて、前記学習データから選別データを生成する第4の工程と、

前記第2の学習済モデル生成部が、該生成された前記選別データとキーワードとを入力し、第2の機械学習を実行して、第2の学習済モデルを生成する第5の工程と、

前記推論結果出力部が、前記第2の学習済モデルとキーワードとを用いて、推論結果としての学習済モデルを出力する第6の工程と、

前記文書データ出力部が、前記第1の学習済モデルを用いて、推論結果としての学習済モデルを評価し、評価の高い文書データを出力する第7の工程と、

を備え、

前記教師データは、前記特定の前記学習データにラベルが付されたデータであり、前記ラベルには、文書に種別に応じた複数のパラメータに対する充足度が示されており、

前記第3の工程において、前記第1の学習済モデル生成部は、前記教師データとして、前記特定の前記学習データとラベルとを入力し、

前記第4の工程において、前記選別データ生成部は、前記第1の学習済モデルを用いて、前記特定の学習データ以外の前記学習データに前記ラベルを付し、前記パラメータごとの充足度が高い前記ラベルが多く付された前記学習データを前記選別データとして生成することを特徴とする文書生成方法。

【請求項8】

学習データ収集部と、抽出部と、第 1 の学習済モデル生成部と、選別データ生成部と、第 2 の学習済モデル生成部と、推論結果出力部と、文書データ出力部と、を含む文書生成システムにおける文書生成方法をコンピュータに実行させるためのプログラムであって、

前記学習データ収集部が、文書データとしての学習データを収集する第 1 の工程と、前記抽出部が、該収集された前記学習データの中から特定の前記学習データを教師データとして抽出する第 2 の工程と、

前記第 1 の学習済モデル生成部が、該抽出された前記教師データを入力し、第 1 の機械学習を実行して、第 1 の学習済モデルを生成する第 3 の工程と、

前記選別データ生成部が、該生成された前記第 1 の学習済モデルを用いて、前記学習データから選別データを生成する第 4 の工程と、

前記第 2 の学習済モデル生成部が、該生成された前記選別データとキーワードとを入力し、第 2 の機械学習を実行して、第 2 の学習済モデルを生成する第 5 の工程と、

前記推論結果出力部が、前記第 2 の学習済モデルとキーワードとを用いて、推論結果としての学習済モデルを出力する第 6 の工程と、

前記文書データ出力部が、前記第 1 の学習済モデルを用いて、推論結果としての学習済モデルを評価し、評価の高い文書データを出力する第 7 の工程と、

を備え、

前記教師データは、前記特定の前記学習データにラベルが付されたデータであり、前記ラベルには、文書に種別に応じた複数のパラメータに対する充足度が示されており、

前記第 3 の工程において、前記第 1 の学習済モデル生成部は、前記教師データとして、前記特定の前記学習データとラベルとを入力し、

前記第 4 の工程において、前記選別データ生成部は、前記第 1 の学習済モデルを用いて、前記特定の学習データ以外の前記学習データに前記ラベルを付し、前記パラメータごとの充足度が高い前記ラベルが多く付された前記学習データを前記選別データとして生成するプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、文書生成システム、文書生成方法およびプログラムに関する。

【背景技術】

【0002】

近年、機械学習に関する研究が盛んに進められており、文書生成に関する分野においても機械学習を用いたシステムが開発されている。

機械学習を用いたシステムでは、推論の対象を特定する要件を定め、予め多数の教師データを用いて、当該推論を行うためのモデルを学習する（例えば、非特許文献 1 参照）。

【0003】

文書生成に関する分野においては、上記の多数の教師データを得る手法として、アノテーションが用いられることが多く、アノテーションは、学習済みモデルの評価に大きな影響を与える重要なプロセスであると言える（例えば、特許文献 1 参照）。

【先行技術文献】

【特許文献】

【0004】

【文献】特開 2021 - 168055 号公報

【非特許文献】

【0005】

【文献】Laura L. Pullum Brian J. Taylor Majorie A. Darrah 「Guidance for the Verification and Validation of Neural Networks」

【0006】

図 7 は、アノテーションと機械学習とを用いた従来の文書生成システムの構成を示した図である。

10

20

30

40

50

図 7 に示すように、文書生成システムは、学習データ収集部 10 と、アノテーション部 20 と、学習済モデル生成部 30 と、学習済モデル記憶部 40 と、推論結果出力部 50 と、を含んで構成され、学習データ収集部 10 と、アノテーション部 20 と、学習済モデル生成部 30 と、学習済モデル記憶部 40 と、により実行される処理工程が学習工程、推論結果出力部 50 と、によって実行される処理工程が推論工程となっている。

【0007】

学習工程では、学習データ収集部 10 は、膨大な学習データを収集し、アノテーション部 20 は、学習データ収集部 10 において収集された膨大な学習データとキーワードとを入力してアノテーション処理を実行する。

アノテーション部 20 において処理された学習データは、学習済モデル生成部 30 において機械学習が実行され、学習済モデルが生成される。

そして、生成された学習済モデルを学習済モデル記憶部 40 に格納して、処理を終了する。

【0008】

推論工程では、推論結果出力部 50 が学習済モデル記憶部 40 から読み出した学習済モデルとキーワードとに基づいて、推論結果を出力する。

【発明の概要】

【発明が解決しようとする課題】

【0009】

一般に、文書生成システムに用いられる学習データは、インターネットや SNS 等に開示された文書データであり、この種の学習データには、ノイズが含まれている。

そのため、こうしたノイズを含んだ文書データを用いて学習済みモデルを生成するために、文書の生成精度が低くなるという課題があった。

【0010】

また、上記の文書の生成精度を抑制するためには、人手を用いて、膨大な学習データの中からノイズを含む学習データを除去する必要があるため、コストが高くなるという課題もあった。

【0011】

そこで、本発明は、上述の課題に鑑みてなされたものであり、コストを抑えつつ、文書生成の精度を向上させる文書生成システム、文書生成方法およびプログラムを提供することを目的とする。

【課題を解決するための手段】

【0012】

形態 1 ; 本発明の 1 またはそれ以上の実施形態は、文書データとしての学習データを収集する学習データ収集部と、該収集された前記学習データの中から特定の前記学習データを教師データとして抽出する抽出部と、該抽出された前記教師データを入力し、第 1 の機械学習を実行して、第 1 の学習済モデルを生成する第 1 の学習済モデル生成部と、該生成された前記第 1 の学習済モデルを用いて、前記学習データから選別データを生成する選別データ生成部と、該生成された前記選別データとキーワードとを入力し、第 2 の機械学習を実行して、第 2 の学習済モデルを生成する第 2 の学習済モデル生成部と、前記第 2 の学習済モデルとキーワードとを用いて、推論結果としての文書データを出力する推論結果出力部と、前記第 1 の学習済モデルを用いて、推論結果としての文書データを評価し、評価の高い文書データを出力する文書データ出力部と、を備えたことを特徴とする文書生成システムを提案している。

【0013】

形態 2 ; 本発明の 1 またはそれ以上の実施形態は、前記教師データは、前記特定の前記学習データにラベルが付された文書データであり、前記ラベルには、文書に種別に応じた複数のパラメータに対する充足度が示されており、前記第 1 の学習済モデル生成部は、前記教師データとして、前記特定の前記学習データとラベルとを入力し、前記選別データ生成部は、前記第 1 の学習済モデルを用いて、前記特定の学習データ以外の前記学習データ

10

20

30

40

50

に前記ラベルを付し、前記パラメータごとの充足度が高い前記ラベルが多く付された前記学習データを選別データとして生成することを特徴とする文書生成システムを提案している。

【0014】

形態3；本発明の1またはそれ以上の実施形態は、前記第1の学習済モデル生成部および前記選別データ生成部は、前記特定の前記学習データと前記複数のパラメータのうち、1つの前記パラメータに対する充足度とを組み合わせた文書データに対して、処理を実行することを特徴とする文書生成システムを提案している。

【0015】

形態4；本発明の1またはそれ以上の実施形態は、前記文書データ出力部から出力される前記文書データは、前記選別データ生成部の前記学習データとしてフィードバックされることを特徴とする文書生成システムを提案している。

10

【0016】

形態5；本発明の1またはそれ以上の実施形態は、前記パラメータに対する充足度は、2値で示されることを特徴とする文書生成システムを提案している。

【0017】

形態6；本発明の1またはそれ以上の実施形態は、前記文書データが求人票の募集要項であることを特徴とする文書生成システムを提案している。

【0018】

形態7；本発明の1またはそれ以上の実施形態は、前記パラメータが、自社の現状、やってほしい仕事、ポジション、仕事の進め方、将来像を含み、前記充足度は、各パラメータに関する記載が前記文書データに記載されているか否かを2値で示していることを特徴とする文書生成システムを提案している。

20

【0019】

形態8；本発明の1またはそれ以上の実施形態は、学習データ収集部と、抽出部と、第1の学習済モデル生成部と、選別データ生成部と、第2の学習済モデル生成部と、推論結果出力部と、文書データ出力部と、を含む文書生成システムにおける文書生成方法であって、前記学習データ収集部が、文書データとしての学習データを収集する第1の工程と、前記抽出部が、該収集された前記学習データの中から特定の前記学習データを教師データとして抽出する第2の工程と、前記第1の学習済モデル生成部が、該抽出された前記教師データを入力し、第1の機械学習を実行して、第1の学習済モデルを生成する第3の工程と、前記選別データ生成部が、該生成された前記第1の学習済モデルを用いて、前記学習データから選別データを生成する第4の工程と、前記第2の学習済モデル生成部が、該生成された前記選別データとキーワードとを入力し、第2の機械学習を実行して、第2の学習済モデルを生成する第5の工程と、前記推論結果出力部が、前記第2の学習済モデルとキーワードとを用いて、推論結果としての学習済モデルを出力する第6の工程と、前記文書データ出力部が、前記第1の学習済モデルを用いて、推論結果としての学習済モデルを評価し、評価の高い文書データを出力する第7の工程と、を備えたことを特徴とする文書生成方法を提案している。

30

【0020】

形態9；本発明の1またはそれ以上の実施形態は、学習データ収集部と、抽出部と、第1の学習済モデル生成部と、選別データ生成部と、第2の学習済モデル生成部と、推論結果出力部と、文書データ出力部と、を含む文書生成システムにおける文書生成方法をコンピュータに実行させるためのプログラムであって、前記学習データ収集部が、文書データとしての学習データを収集する第1の工程と、前記抽出部が、該収集された前記学習データの中から特定の前記学習データを教師データとして抽出する第2の工程と、前記第1の学習済モデル生成部が、該抽出された前記教師データを入力し、第1の機械学習を実行して、第1の学習済モデルを生成する第3の工程と、前記選別データ生成部が、該生成された前記第1の学習済モデルを用いて、前記学習データから選別データを生成する第4の工程と、前記第2の学習済モデル生成部が、該生成された前記選別データとキーワードとを

40

50

入力し、第 2 の機械学習を実行して、第 2 の学習済モデルを生成する第 5 の工程と、前記推論結果出力部が、前記第 2 の学習済モデルとキーワードとを用いて、推論結果としての学習済モデルを出力する第 6 の工程と、前記文書データ出力部が、前記第 1 の学習済モデルを用いて、推論結果としての学習済モデルを評価し、評価の高い文書データを出力する第 7 の工程と、をコンピュータに実行させるためのプログラムを提案している。

【発明の効果】

【0021】

本発明の 1 またはそれ以上の実施形態によれば、コストを抑えつつ、文書生成の精度を向上させることができる。

【図面の簡単な説明】

10

【0022】

【図 1】本発明の第 1 の実施形態に係る文書生成システムの構成を示す図である。

【図 2】本発明の第 1 の実施形態に係る文書生成システムの処理を示す図である。

【図 3】本発明の第 2 の実施形態に係る文書生成システムの構成を示す図である。

【図 4】本発明の第 2 の実施形態に係る文書生成システムの処理を示す図である。

【図 5】本発明の第 3 の実施形態に係る文書生成システムの構成を示す図である。

【図 6】本発明の第 3 の実施形態に係る文書生成システムの処理を示す図である。

【図 7】従来例に係る文書生成システムの構成を示す図である。

【発明を実施するための形態】

【0023】

20

以下、図 1 から図 6 を用いて、本発明の実施形態について、説明する。

【0024】

< 第 1 の実施形態 >

以下、図 1、図 2 を用いて、本実施形態に係る文書生成システム 1 について、説明する。

【0025】

< 文書生成システム 1 の構成 >

図 1 に示すように、本実施形態に係る文書生成システム 1 は、学習データ収集部 110 と、抽出部 120 と、第 1 の学習済モデル生成部 130 と、第 1 の学習済モデル記憶部 140 と、選別データ生成部 150 と、第 2 の学習済モデル生成部 160 と、第 2 の学習済モデル記憶部 170 と、推論結果出力部 180 と、文書データ出力部 190 と、を含んで構成されている。

30

【0026】

学習データ収集部 110 は、文書データとしての学習データを収集する。

学習データ収集部 110 において収集された学習データは、後述する抽出部 120 に出力される。

なお、文書データとしては、例えば、報告書やプレゼン資料等のビジネス文書をはじめとする種々の文書を例示できるが、以下では、求人票の募集要項を文書データとして例示して、説明する。

【0027】

抽出部 120 は、学習データ収集部 110 において収集された学習データの中から特定の学習データを教師データとして抽出する。

40

抽出部 120 において抽出された教師データは、後述する第 1 の学習済モデル生成部 130 に出力される。

また、抽出部 120 において抽出された教師データは、図示しない記憶部に記憶保存される。

教師データは、例えば、特定の学習データにラベルが付された文書データであり、ラベルには、文書に種別に応じた複数のパラメータに対する充足度が示されている。

また、パラメータに対する充足度は、例えば、2 値で示される。

例えば、文書データを求人票の募集要項とした場合、パラメータとしては、自社の現状、やってほしい仕事、ポジション、仕事の進め方、将来像等を例示でき、充足度は、各パ

50

ラメータに関する記載が文書データに記載されているか否かを2値で示している。

抽出部120は、学習データ収集部110において収集された学習データの中から、複数のパラメータに対して、その充足度が少なくとも、1つ満たされた厳選された少量の教師データを抽出する。

【0028】

第1の学習済モデル生成部130は、抽出された教師データを入力し、第1の機械学習を実行して、第1の学習済モデルを生成する。

ここで、第1の学習済モデル生成部130において生成される第1の学習済モデルは、後述する選別データ生成部において、選別データを生成するために用いられる学習済モデルである。

第1の学習済モデル生成部130において生成された第1の学習済モデルは、後述する第1の学習済モデル記憶部140に記憶される。

【0029】

第1の学習済モデル記憶部140は、RAM(Random Access Memory)等から構成され、第1の学習済モデルを記憶保存する。

【0030】

選別データ生成部150は、第1の学習済モデル記憶部140から第1の学習済モデルを読み出して、学習データから選別された選別データを生成する。

選別データ生成部150において生成された選別データは、膨大な量の文書データとなる。

選別データ生成部150において生成された選別データは、後述する第2の学習済モデル生成部160に出力される。

【0031】

第2の学習済モデル生成部160は、選別データ生成部150において生成された選別データとキーワードとを入力し、第2の機械学習を実行して、第2の学習済モデルを生成する。

キーワードは文書の種別や内容に応じたものであり、予め準備されたものである。

第2の学習済モデル生成部160において生成された第2の学習済モデルは、後述する第2の学習済モデル記憶部170に記憶される。

【0032】

第2の学習済モデル記憶部170は、RAM(Random Access Memory)等から構成され、第2の学習済モデルを記憶保存する。

【0033】

推論結果出力部180は、第2の学習済モデル生成部160において生成された第2の学習済モデルとキーワードとを用いて、推論結果としての文書データを出力する。

推論結果出力部180における推論結果としての文書データは、後述する文書データ出力部190に出力される。

【0034】

文書データ出力部190は、第1の学習済モデル生成部130において生成された第1の学習済モデルを用いて、推論結果としての文書データを評価し、評価の高い文書データを出力する。

なお、文書データ出力部190の評価としては、ラベルに示された文書に種別に応じた複数のパラメータに対する充足度が高いもの、具体的には、充足度の高いパラメータの数により、評価を行う。

【0035】

<文書生成システム1の処理>

図2を用いて、本実施形態に係る文書生成システム1の処理について説明する。

【0036】

図2に示すように、学習データ収集部110は、文書データとしての学習データを収集する(ステップS110)。

10

20

30

40

50

学習データ収集部 110 において収集された学習データは、抽出部 120 に出力される。

【0037】

抽出部 120 は、学習データ収集部 110 において収集された学習データの中から特定の学習データを教師データとして抽出する（ステップ S120）。

抽出部 120 において抽出された教師データは、第 1 の学習済モデル生成部 130 に出力される。

【0038】

第 1 の学習済モデル生成部 130 は、抽出された教師データを入力し、第 1 の機械学習を実行して、第 1 の学習済モデルを生成する（ステップ S130）。

第 1 の学習済モデル生成部 130 において生成された第 1 の学習済モデルは、第 1 の学習済モデル記憶部 140 に記憶される。

【0039】

選別データ生成部 150 は、第 1 の学習済モデル記憶部 140 から第 1 の学習済モデルを読み出して、学習データから選別した選別データを生成する（ステップ S140）。

選別データ生成部 150 において生成された膨大な量の選別データは、第 2 の学習済モデル生成部 160 に出力される。

【0040】

第 2 の学習済モデル生成部 160 は、選別データ生成部 150 において生成された選別データとキーワードとを入力し、第 2 の機械学習を実行して、第 2 の学習済モデルを生成する（ステップ S150）。

第 2 の学習済モデル生成部 160 において生成された第 2 の学習済モデルは、第 2 の学習済モデル記憶部 170 に記憶される。

【0041】

推論結果出力部 180 は、第 2 の学習済モデル生成部 160 において生成された第 2 の学習済モデルとキーワードとを用いて、推論結果としての文書データを出力する（ステップ S160）。

推論結果出力部 180 における推論結果としての文書データは、文書データ出力部 190 に出力される。

【0042】

文書データ出力部 190 は、第 1 の学習済モデル生成部 130 において生成された第 1 の学習済モデルを用いて、推論結果としての文書データを評価し、評価の高い文書データを出力する（ステップ S160）。

【0043】

< 作用・効果 >

以上、説明したように、本実施形態に係る文書生成システム 1 は、文書データとしての学習データを収集する学習データ収集部 110 と、該収集された学習データの中から特定の学習データを教師データとして抽出する抽出部 120 と、該抽出された教師データを入力し、第 1 の機械学習を実行して、第 1 の学習済モデルを生成する第 1 の学習済モデル生成部 130 と、該生成された第 1 の学習済モデルを用いて、学習データから選別した文書データを生成する選別データ生成部 150 と、該生成された選別データとキーワードとを入力し、第 2 の機械学習を実行して、第 2 の学習済モデルを生成する第 2 の学習済モデル生成部 160 と、第 2 の学習済モデルとキーワードとを用いて、推論結果としての文書データを出力する推論結果出力部 180 と、第 1 の学習済モデルを用いて、推論結果としての文書データを評価し、評価の高い文書データを出力する文書データ出力部 190 と、を備えている。

抽出部 120 は、収集された学習データの中から特定の学習データを教師データとして抽出する。

ここで、教師データは、例えば、特定の学習データにラベルが付された文書データであり、ラベルには、文書に種別に応じた複数のパラメータに対する充足度が示されている。

抽出部 120 は、学習データ収集部 110 において収集された学習データの中から、複

10

20

30

40

50

数のパラメータに対して、その充足度が少なくとも１つ満たされた厳選された少量の教師データを抽出する。

つまり、抽出部１２０は、学習データ収集部１１０において収集された学習データの中から、複数のパラメータに対して、その充足度が少なくとも１つ満たされた厳選された少量の教師データを抽出するため、工数を大幅に削減できる。

なお、これが実現できるのは、後述する第１の学習済モデル生成部１３０、選別データ生成部１５０、第２の学習済モデル生成部１６０、推論結果出力部１８０、文書データ出力部１９０を備えているためである。

第１の学習済モデル生成部１３０は、抽出部１２０において抽出された教師データを入力し、第１の機械学習を実行して、第１の学習済モデルを生成し、選別データ生成部１５０は、該生成された第１の学習済モデルを用いて、学習データから選別した文書データを生成し、第２の学習済モデル生成部１６０は、該生成された選別データとキーワードとを入力し、第２の機械学習を実行して、第２の学習済モデルを生成する。

10

つまり、第１の学習済モデル生成部１３０は、教師データを入力し、第１の学習済モデルを生成し、選別データ生成部１５０は、第１の学習済モデルを用いて、学習データから選別した膨大な選別データを生成し、第２の学習済モデル生成部１６０は、選別データとキーワードとを入力し、第２の学習済モデルを生成する。

そのため、厳選された少量の教師データから第１の学習済モデルが生成することによって、精度の高い学習済モデルを生成することができる。

また、第１の学習済モデルを用いて、学習データから選別した膨大な文書データが生成されるため、少ない工数で、精度が高く、従来のアノテーション処理を凌駕する文書データを生成することができる。

20

さらに、選別データとキーワードとを入力し、第２の学習済モデルが生成されるため、精度の高い処理が実現できる。

推論結果出力部１８０は、第２の学習済モデルとキーワードとを用いて、推論結果としての文書データを出力し、文書データ出力部１９０は、第１の学習済モデルを用いて、推論結果としての文書データを評価し、評価の高い文書データを出力する。

つまり、推論結果出力部１８０は、第１の学習済モデルを用いて、学習データから選別した膨大な選別データから生成された第２の学習済モデルから推論結果としての文書データを出力し、文書データ出力部１９０は、第１の学習済モデル生成部１３０において生成された第１の学習済モデルを用いて、推論結果としての文書データを評価し、評価の高い文書データを出力するため、確度の高い推論と評価によって、高い品質の文書データを生成することができる。

30

【００４４】

< 第２の実施形態 >

以下、図３、図４を用いて、本実施形態に係る文書生成システム１Ａについて、説明する。

【００４５】

< 文書生成システム１Ａの構成 >

図３に示すように、本実施形態に係る文書生成システム１Ａは、学習データ収集部１１０と、抽出部１２０Ａと、第１の学習済モデル生成部１３０Ａと、第１の学習済モデル記憶部１４０Ａと、選別データ生成部１５０Ａと、第２の学習済モデル生成部１６０と、第２の学習済モデル記憶部１７０と、推論結果出力部１８０と、文書データ出力部１９０と、を含んで構成されている。

40

なお、第１の実施形態と同一の符号を付す構成要素については、同様の機能を有することから、その詳細な説明は、省略する。

【００４６】

抽出部１２０Ａは、学習データ収集部１１０において収集された学習データの中から特定の学習データを教師データとして抽出する。

抽出部１２０Ａは、抽出された教師データを学習データとラベルとに分離して、後述す

50

る第1の学習済モデル生成部130Aに出力する。

【0047】

第1の学習済モデル生成部130Aは、教師データとしての学習データとラベルとを入力し、第1の機械学習を実行して、第1の学習済モデルを生成する。

第1の学習済モデル生成部130において生成された第1の学習済モデルは、後述する第1の学習済モデル記憶部140Aに記憶される。

なお、第1の学習済モデル生成部130Aは、特定の学習データと複数のパラメータのうち、1つのパラメータに対する充足度とを組み合わせた文書データに対して、繰り返し処理を実行する。

【0048】

第1の学習済モデル記憶部140Aは、RAM(Random Access Memory)等から構成され、第1の学習済モデル生成部130Aにおいて生成された第1の学習済モデルを記憶保存する。

【0049】

選別データ生成部150Aは、第1の学習済モデル記憶部140Aから第1の学習済モデルを読み出して、特定の学習データ以外の学習データにラベルを付し、パラメータごとの充足度が高い学習データを選別データとして生成する。

選別データ生成部150Aにおいて生成される選別データは、膨大な量の文書データとなる。

選別データ生成部150Aにおいて生成された選別データは、後述する第2の学習済モデル生成部160に出力される。

なお、選別データ生成部150Aは、特定の学習データと複数のパラメータのうち、1つのパラメータに対する充足度とを組み合わせた文書データに対して、繰り返し処理を実行する。

【0050】

<文書生成システム1の処理>

図2を用いて、本実施形態に係る文書生成システム1の処理について説明する。

【0051】

図2に示すように、学習データ収集部110は、文書データとしての学習データを収集する(ステップS110)。

学習データ収集部110において収集された学習データは、抽出部120に出力される。

【0052】

抽出部120Aは、学習データ収集部110において収集された学習データの中から特定の学習データを教師データとして抽出する(ステップS210)。

抽出部120Aは、抽出された教師データを学習データとラベルとに分離して、後述する第1の学習済モデル生成部130Aに出力する。

【0053】

第1の学習済モデル生成部130Aは、教師データとしての学習データとラベルとを入力し、第1の機械学習を実行して、第1の学習済モデルを生成する(ステップS220)。

なお、第1の学習済モデル生成部130Aは、特定の学習データと複数のパラメータのうち、1つのパラメータに対する充足度とを組み合わせた文書データに対して、繰り返し処理を実行する。

第1の学習済モデル生成部130Aにおいて生成された第1の学習済モデルは、第1の学習済モデル記憶部140Aに記憶される。

【0054】

選別データ生成部150Aは、第1の学習済モデル記憶部140Aから第1の学習済モデルを読み出して、特定の学習データ以外の学習データにラベルを付し、パラメータごとの充足度が高い学習データを選別データとして生成する(ステップS230)。

選別データ生成部150Aにおいて生成された膨大な量の選別データは、第2の学習済モデル生成部160に出力される。

10

20

30

40

50

なお、選別データ生成部 150A は、特定の学習データと複数のパラメータのうち、1つのパラメータに対する充足度とを組み合わせた文書データに対して、繰り返し処理を実行する。

【0055】

第2の学習済モデル生成部 160 は、選別データ生成部 150A において生成された選別データとキーワードとを入力し、第2の機械学習を実行して、第2の学習済モデルを生成する（ステップ S150）。

第2の学習済モデル生成部 160 において生成された第2の学習済モデルは、第2の学習済モデル記憶部 170 に記憶される。

【0056】

推論結果出力部 180 は、第2の学習済モデル生成部 160 において生成された第2の学習済モデルとキーワードとを用いて、推論結果としての文書データを出力する（ステップ S160）。

推論結果出力部 180 における推論結果としての文書データは、文書データ出力部 190 に出力される。

【0057】

文書データ出力部 190 は、第1の学習済モデル生成部 130 において生成された第1の学習済モデルを用いて、推論結果としての文書データを評価し、評価の高い文書データを出力する（ステップ S170）。

【0058】

<作用・効果>

以上、説明したように、本実施形態に係る文書生成システム 1A において、教師データは、特定の学習データにラベルが付された文書データであり、ラベルには、文書の種別に応じた複数のパラメータに対する充足度が示されており、第1の学習済モデル生成部 130A は、教師データとして、特定の学習データとラベルとを入力し、選別データ生成部 150A は、第1の学習済モデルを用いて、特定の学習データ以外の学習データにラベルを付し、パラメータごとの充足度が高い学習データを選別データとして生成する。

つまり、第1の学習済モデル生成部 130A は、教師データとして、特定の学習データとラベルとを入力し、選別データ生成部 150A は、第1の学習済モデルを用いて、特定の学習データ以外の学習データにラベルを付し、パラメータごとの充足度が高い学習データを選別データとして生成する。

そのため、パラメータごとの充足度が高い学習データを選別データとして生成することにより、教師データが従来よりも少数であっても、文書生成の精度を向上させることができる。

また、教師データが従来よりも少数であることから、教師データを得るための工数が削減され、システム全体のコストを低減することができる。

【0059】

また、本実施形態に係る文書生成システム 1 において、第1の学習済モデル生成部 130A および選別データ生成部 150A は、特定の学習データと複数のパラメータのうち、1つのパラメータに対する充足度とを組み合わせた文書データに対して、繰り返し処理を実行する。

つまり、第1の機械学習により、ラベルが付された学習データを生成し、その中からパラメータごとの充足度が高い学習データを選別データとして生成する。

そのため、パラメータごとの充足度が高い学習データを選別データとして生成することにより、教師データが従来よりも少数であっても、文書生成の精度を向上させることができる。

【0060】

<第3の実施形態>

以下、図5、図6を用いて、本実施形態に係る文書生成システム 1B について、説明する。

10

20

30

40

50

【 0 0 6 1 】**< 文書生成システム 1 B の構成 >**

図 5 に示すように、本実施形態に係る文書生成システム 1 B は、学習データ収集部 1 1 0 と、抽出部 1 2 0 A と、第 1 の学習済モデル生成部 1 3 0 A と、第 1 の学習済モデル記憶部 1 4 0 A と、選別データ生成部 1 5 0 A と、第 2 の学習済モデル生成部 1 6 0 と、第 2 の学習済モデル記憶部 1 7 0 と、推論結果出力部 1 8 0 と、文書データ出力部 1 9 0 A と、を含んで構成されている。

なお、第 1 の実施形態および第 2 の実施形態と同一の符号を付す構成要素については、同様の機能を有することから、その詳細な説明は、省略する。

【 0 0 6 2 】

文書データ出力部 1 9 0 A は、第 1 の学習済モデル生成部 1 3 0 において生成された第 1 の学習済モデルを用いて、推論結果としての文書データを評価し、評価の高い文書データを出力する。

また、文書データ出力部 1 9 0 A から出力される評価の高い文書データは、選別データ生成部 1 5 0 A の学習データとしてフィードバックされる。

【 0 0 6 3 】**< 文書生成システム 1 B の処理 >**

図 6 を用いて、本実施形態に係る文書生成システム 1 B の処理について説明する。

【 0 0 6 4 】

図 2 に示すように、学習データ収集部 1 1 0 は、文書データとしての学習データを収集する（ステップ S 1 1 0 ）。

学習データ収集部 1 1 0 において収集された学習データは、抽出部 1 2 0 に出力される。

【 0 0 6 5 】

抽出部 1 2 0 A は、学習データ収集部 1 1 0 において収集された学習データの中から特定の学習データを教師データとして抽出する（ステップ S 2 1 0 ）。

抽出部 1 2 0 A は、抽出された教師データを学習データとラベルとに分離して、後述する第 1 の学習済モデル生成部 1 3 0 A に出力する。

【 0 0 6 6 】

第 1 の学習済モデル生成部 1 3 0 A は、教師データとしての学習データとラベルとを入力し、第 1 の機械学習を実行して、第 1 の学習済モデルを生成する（ステップ S 2 2 0 ）。

第 1 の学習済モデル生成部 1 3 0 A において生成された第 1 の学習済モデルは、第 1 の学習済モデル記憶部 1 4 0 A に記憶される。

なお、第 1 の学習済モデル生成部 1 3 0 A は、特定の学習データと複数のパラメータのうち、1 つのパラメータに対する充足度とを組み合わせた文書データに対して、繰り返し処理を実行する。

【 0 0 6 7 】

選別データ生成部 1 5 0 A は、第 1 の学習済モデル記憶部 1 4 0 A から第 1 の学習済モデルを読み出して、特定の学習データ以外の学習データにラベルを付し、パラメータごとの充足度が高いラベルが多く付された学習データを選別データとして生成する（ステップ S 2 3 0 ）。

選別データ生成部 1 5 0 A において生成された膨大な量の選別データは、第 2 の学習済モデル生成部 1 6 0 に出力される。

なお、選別データ生成部 1 5 0 A は、特定の学習データと複数のパラメータのうち、1 つのパラメータに対する充足度とを組み合わせた文書データに対して、繰り返し処理を実行する。

【 0 0 6 8 】

第 2 の学習済モデル生成部 1 6 0 は、選別データ生成部 1 5 0 A において生成された選別データとキーワードとを入力し、第 2 の機械学習を実行して、第 2 の学習済モデルを生成する（ステップ S 1 5 0 ）。

第 2 の学習済モデル生成部 1 6 0 において生成された第 2 の学習済モデルは、第 2 の学

10

20

30

40

50

習済モデル記憶部 170 に記憶される。

【0069】

推論結果出力部 180 は、第 2 の学習済モデル生成部 160 において生成された第 2 の学習済モデルとキーワードとを用いて、推論結果としての文書データを出力する（ステップ S160）。

推論結果出力部 180 における推論結果としての文書データは、文書データ出力部 190 に出力される。

【0070】

文書データ出力部 190 A は、第 1 の学習済モデル生成部 130 において生成された第 1 の学習済モデルを用いて、推論結果としての文書データを評価し、評価の高い文書データを出力する。

また、文書データ出力部 190 A から出力される文書データは、選別データ生成部 150 A の学習データとしてフィードバックされる。

【0071】

<作用・効果>

以上、説明したように、本実施形態に係る文書生成システム 1 B において、文書データ出力部 190 A から出力される文書データは、選別データ生成部 150 A の学習データとしてフィードバックされる。

つまり、文書データ出力部 190 A において、評価が高いとされる文書データを選別データ生成部 150 A で用いられる学習データに加える。

そのため、更に精度の高い文書データの作成が期待できる。

【0072】

<応用例>

本実施形態に係る文書生成システム 1、1 A、1 B を用いることによって、Web サイトや求人サイトあるいは報告書、プレゼン資料等の種々の文書を自動的に生成することができる。

また、文書生成の観点やポイントを教育するためのツールとして活用することも期待できる。

また、文書を添削するためのツールや記載されている文書の構成や内容を分析するためのツールとして活用することも期待できる。

【0073】

なお、第 1 の学習済モデル生成部 130、130 A、選別データ生成部 150、150 A、第 2 の学習済モデル生成部 160、推論結果出力部 180、文書データ出力部 190、190 A の処理をコンピュータシステムが読み取り可能な記録媒体に記録し、この記録媒体に記録されたプログラムを第 1 の学習済モデル生成部 130、130 A、選別データ生成部 150、150 A、第 2 の学習済モデル生成部 160、推論結果出力部 180、文書データ出力部 190、190 A に読み込ませ、実行することによって本発明の文書生成システム 1、1 A、1 B を実現することができる。

ここでいうコンピュータシステムとは、OS や周辺装置等のハードウェアを含む。

【0074】

また、「コンピュータシステム」は、WWW (World Wide Web) システムを利用している場合であれば、ホームページ提供環境（あるいは表示環境）も含むものとする。また、上記プログラムは、このプログラムを記憶装置等に格納したコンピュータシステムから、伝送媒体を介して、あるいは、伝送媒体中の伝送波により他のコンピュータシステムに伝送されてもよい。ここで、プログラムを伝送する「伝送媒体」は、インターネット等のネットワーク（通信網）や電話回線等の通信回線（通信線）のように情報を伝送する機能を有する媒体のことをいう。

【0075】

また、上記プログラムは、前述した機能の一部を実現するためのものであってもよい。さらに、前述した機能をコンピュータシステムにすでに記録されているプログラムとの組

10

20

30

40

50

合せて実現できるもの、いわゆる差分ファイル（差分プログラム）であってもよい。

【 0 0 7 6 】

以上、この発明の実施形態につき、図面を参照して詳述してきたが、具体的な構成はこの実施形態に限られるものではなく、この発明の要旨を逸脱しない範囲の設計等も含まれる。

【 符号の説明 】

【 0 0 7 7 】

1 ; 文書データ生成システム	
1 A ; 文書データ生成システム	
1 B ; 文書データ生成システム	10
1 1 0 ; 学習データ収集部	
1 2 0 ; 抽出部	
1 2 0 A ; 抽出部	
1 3 0 ; 第 1 の学習済モデル生成部	
1 3 0 A ; 第 1 の学習済モデル生成部	
1 4 0 ; 第 1 の学習済モデル記憶部	
1 4 0 A ; 第 1 の学習済モデル記憶部	
1 5 0 ; 選別データ生成部	
1 5 0 A ; 選別データ生成部	
1 6 0 ; 第 2 の学習済モデル生成部	20
1 7 0 ; 第 2 の学習済モデル記憶部	
1 8 0 ; 推論結果出力部	
1 9 0 ; 文書データ出力部	
1 9 0 A ; 文書データ出力部	

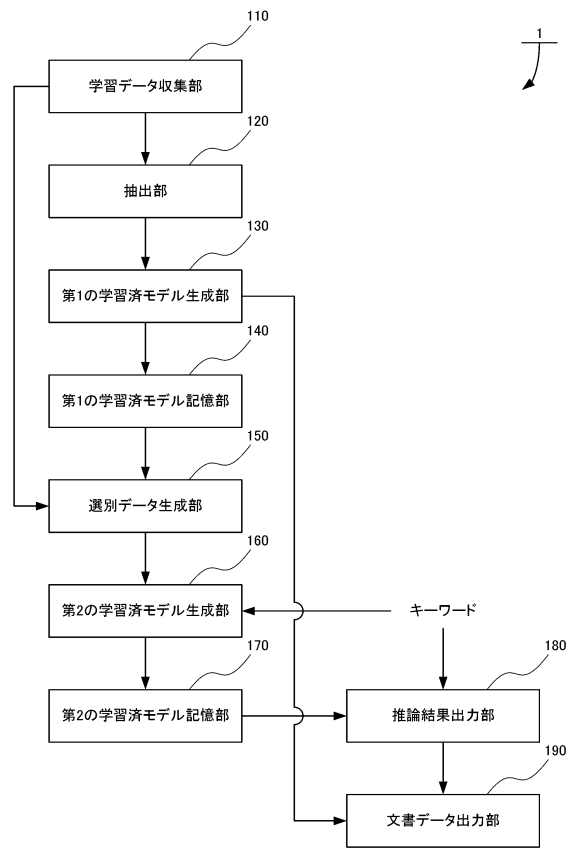
30

40

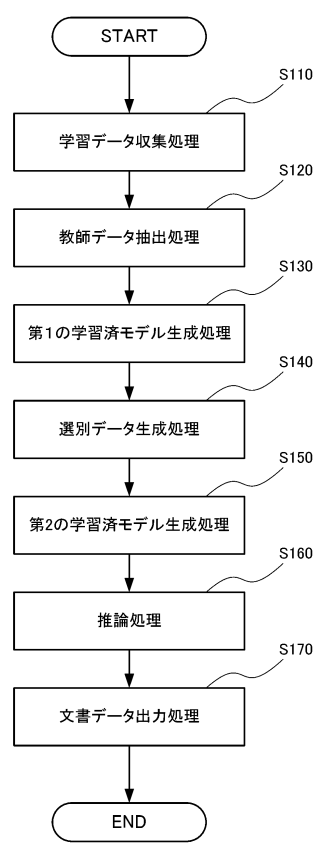
50

【図面】

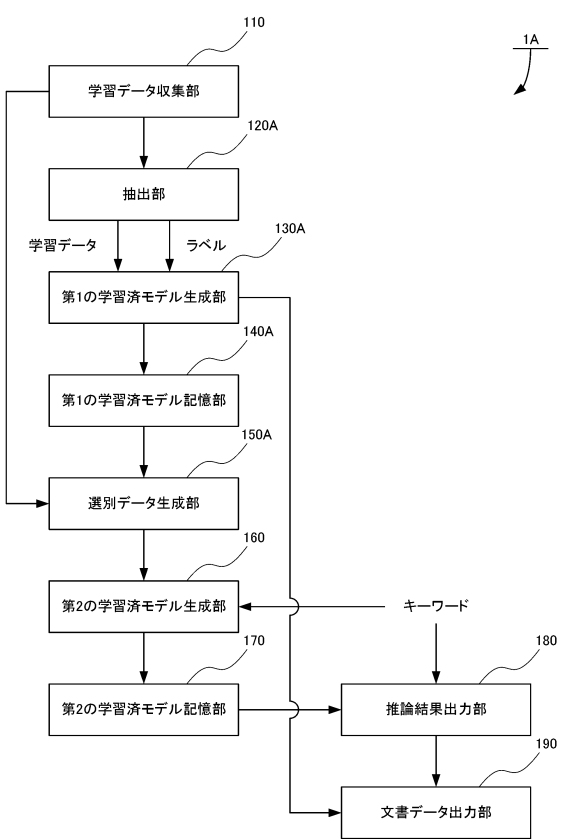
【図 1】



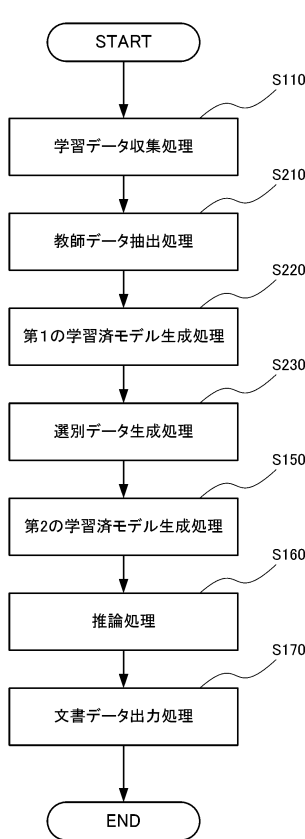
【図 2】



【図 3】



【図 4】



10

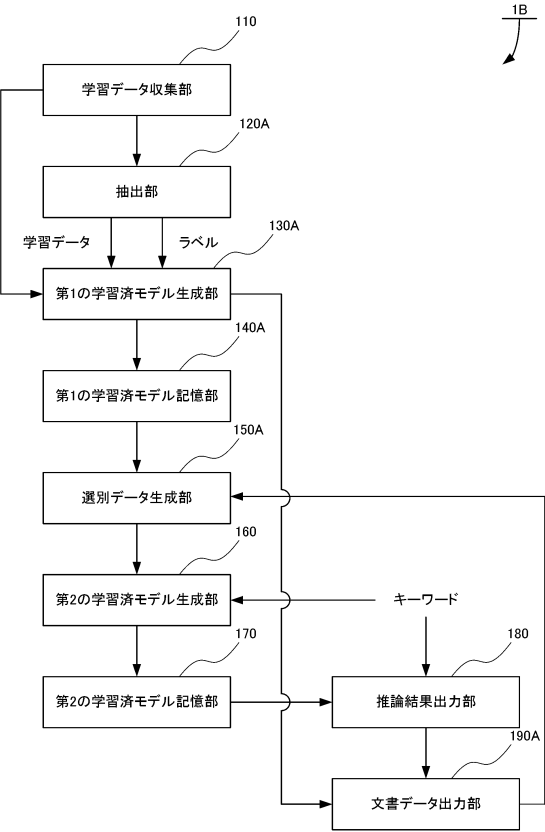
20

30

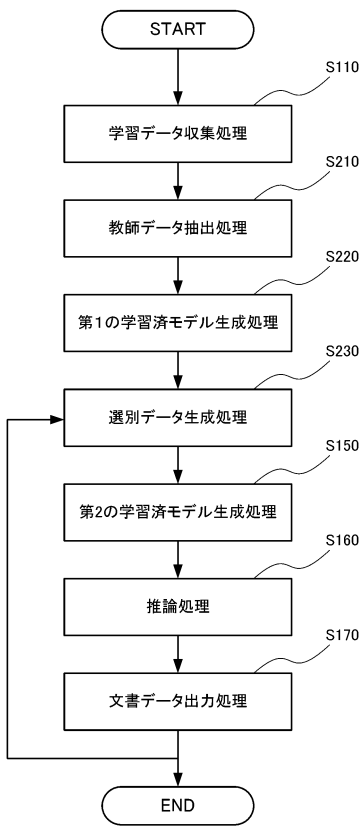
40

50

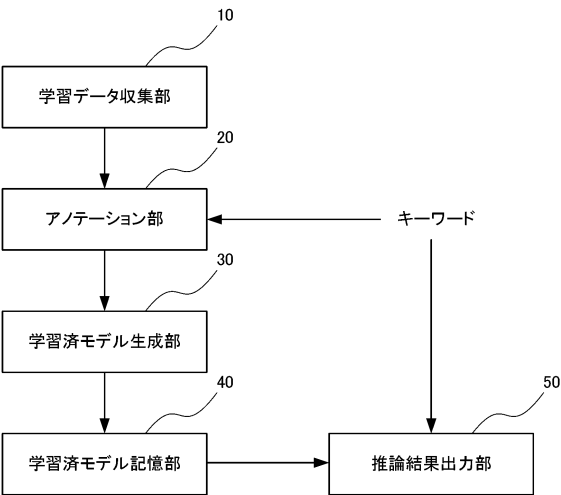
【図 5】



【図 6】



【図 7】



10

20

30

40

50

フロントページの続き

- (56)参考文献 特開 2 0 2 1 - 1 7 9 8 5 9 (J P , A)
 特開 2 0 1 9 - 1 5 9 5 7 6 (J P , A)
 特開 2 0 2 0 - 1 1 9 0 4 4 (J P , A)
 国際公開第 2 0 1 4 / 1 3 6 3 1 6 (W O , A 1)
- (58)調査した分野 (Int.Cl. , D B 名)
- G 0 6 N 3 / 0 9
 G 0 6 N 2 0 / 0 0