(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2023年12月7日(07.12.2023)

WIPO | PCT

(10) 国際公開番号
WO 2023/233633 A1

(51) 国際特許分類:
G06F 40/166 (2020.01) *G06F 40/253* (2020.01)
G06F 40/216 (2020.01) *G06F 40/44* (2020.01)
G06F 40/232 (2020.01)

(21) 国際出願番号： PCT/JP2022/022525

(22) 国際出願日： 2022年6月2日(02.06.2022)

(25) 国際出願の言語： 日本語

(26) 国際公開の言語： 日本語

(71) 出願人: 富士通株式会社 (**FUJITSU LIMITED**) [JP/JP]; 〒2118588 神奈川県川崎市中原区上小田中４丁目１番１号 Kanagawa (JP).

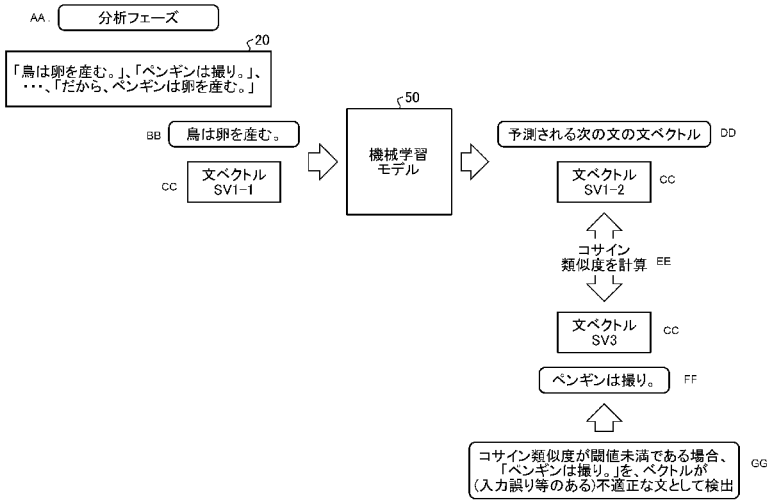
(72) 発明者: 片岡 正弘 (**KATAOKA, Masahiro**); 〒2118588 神奈川県川崎市中原区上小田中４丁目１番１号 富士通株式会社内 Kanagawa (JP). 松村 量 (**MATSUMURA, Ryo**); 〒2118588 神奈川県川崎市中原区上小田中４丁目１番１号 富士通株式会社内 Kanagawa (JP). 尾上 聡 (**ONOUE, Satoshi**); 〒2118588 神奈川県川崎市中原区上小田中４丁目１番１号 富士通株式会社内 Kanagawa (JP).

(74) 代理人: 弁理士法人酒井国際特許事務所 (**SAKAI INTERNATIONAL PATENT OFFICE**); 〒1000013 東京都千代田区霞が関３丁目８番１号 虎の門三井ビルディング Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN,

(54) **Title:** INFORMATION PROCESSING PROGRAM, INFORMATION PROCESSING METHOD, AND INFORMATION PROCESSING DEVICE

(54) 発明の名称： 情報処理プログラム、情報処理方法および情報処理装置



20 "Bird lays egg.", "Penguin takes photo.", "Therefore, penguin lays egg."
50 Machine-learning model
AA Analysis phase
BB Bird lays egg
CC Sentence vectors
DD Sentence vectors of predicted next sentence
EE Compute cosine similarity
FF Penguin takes photo.
GG When cosine similarity is smaller than threshold, detect "Penguin takes photo."as sentence of improper vectors (having input error, etc.)

(57) **Abstract:** This information processing device calculates vectors of a plurality of sentences having a relationship in front and back sentences in a plurality of continuous sentences. The information processing device generates a machine-learning model which predicts sentence vectors of a sentence that is input after a certain sentence when the sentence vectors of the certain sentence are input to the machine-learning model by inputting, in order, the plurality of sentence vectors to the machine-learning model and training the machine-learning model. The information processing device cal-

WO 2023/233633 A1

［続葉有］

CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一　国際調査報告（条約第21条(3)）

culates vectors of a first sentence and vectors of a second sentence that follows the first sentence. The information processing device calculates vectors of a sentence that is predicted to follow the first sentence by inputting the vectors of the first sentence to the machine-learning model, and determines whether the vectors of the second sentence are proper.

(57) 要約：情報処理装置は、連続する複数の文であって、前後の文に関係性を有する複数の文のベクトルをそれぞれ算出する。情報処理装置は、複数の文のベクトルを順番に機械学習モデルに入力して訓練することで、機械学習モデルにある文のベクトルを入力した際に、ある文の次に入力される文の文ベクトルを予測する機械学習モデルを生成する。情報処理装置は、第１文のベクトルと、第１文に続く第２文のベクトルとを算出する。情報処理装置は、第１文のベクトルを機械学習モデルに入力することで第１文に続くと予測される文のベクトルを算出し、第２文のベクトルが適正であるか否かを判定する。

DESCRIPTION

Title of Invention

INFORMATION PROCESSING PROGRAM, INFORMATION PROCESSING METHOD, AND INFORMATION PROCESSING APPARATUS

Technical Field

[0001]

The present invention relates to an information processing program, and the like.

Background Art

[0002]

In recent years, services for calculating vectors of sentences and executing various types of processing such as translation into another language and database search by using the calculated vectors are provided. However, when there is an input error or the like in a sentence itself designated by a user, the vectors of the sentences may not be accurately calculated, and an error may occur in processing such as translation and search.

[0003]

For example, an appropriate sentence "the function is a feature" and an input error sentence "the yesterday is a feature" are sentences having significantly different meanings from each other, and vectors of the sentences are also significantly different from each other.

[0004]

As a related art for revising an input error of a sentence, there is a related art in which an input error of a target sentence is revised by training a learning model by using a data set of pairs of an input error and a corresponding revised sentence from a revision history and inputting the target sentence to the trained learning model.

Citation List

Patent Literature

[0005]

PTL 1: Japanese Laid-open Patent Publication No. 2019-101993

Non Patent Literature

[0006]

NPL 1: Kazuhiro Miki et al., *"Answering English Fill-in-the-blank Questions Using BERT"*, Department of Information Technology Faculty of Engineering, Okayama University, DEIM2020 G2-4 (day1 p47)

NPL 2: Tanaka Yu et al., *'"Building Japanese Input Error Dataset Using Wikipedias Revision History"*, Department of Intelligence Science and Technology, Kyoto University Graduate School, The Association for Natural Language Processing, 26-th Annual Meeting, Journal, March 2020.

Summary of Invention

Technical Problem

[0007]

In the related art described above, there is a technology of filling a blank in which a word or the like of a part of a text is masked, and this technology has high accuracy of a word for filling a blank in a sentence constituted by a plurality of words. However, there are few technologies of increasing the accuracy of a sentence for filling a blank in a text constituted by a plurality of sentences, and there is no technology of detecting a sentence having an input error. In addition, in the related art, an input error such as a typographical error or an omission may be revised, but an input error due to incorrect conversion may not be revised in many cases.

[0008]

In one aspect, an object of the present invention is to provide a an information processing program, an information processing method, and an information processing apparatus that are capable of estimating a sentence to be filled in a blank in a text constituted by a plurality of sentences and detecting a

sentence having an input error.

Solution to Problem

[0009]

In a first aspect, a computer is caused to execute the following processing. The computer calculates individual vectors of a plurality of continuous sentences that have a relationship with preceding and following sentences. The computer generates a machine learning model that predicts a sentence vector of a sentence input next to a certain sentence when a vector of the certain sentence is input to the machine learning model, by sequentially inputting the vectors of the plurality of sentences to the machine learning model and training the machine learning model. The computer calculates a vector of a first sentence and a vector of a second sentence next to the first sentence. The computer calculates a vector of a sentence predicted to be next to the first sentence by inputting the vector of the first sentence to the machine learning model, and determines whether or not the vector of the second sentence is appropriate.

Advantageous Effects of Invention

[0010]

The sentence for filling the blank in the text constituted by the plurality of sentences may be estimated, and the sentence having the input error may be detected.

Brief Description of Drawings

[0011]

[FIG. 1] FIG. 1 is a diagram for describing processing in a learning phase of an information processing apparatus according to the present embodiment.

[FIG. 2] FIG. 2 is a diagram for describing processing in an analysis phase of the information processing apparatus according to the present embodiment.

[FIG. 3] FIG. 3 is a functional block diagram illustrating a configuration of the information processing apparatus according to the present embodiment.

[FIG. 4] FIG. 4 is a diagram illustrating an example of a data structure of a

word vector dictionary.

[FIG. 5A] FIG. 5A is a diagram (1) for describing processing of calculating a sentence vector.

[FIG. 5B] FIG. 5B is a diagram (2) for describing the processing of calculating the sentence vector.

[FIG. 6] FIG. 6 is a diagram for describing processing of generating a sentence transposition index.

[FIG. 7] FIG. 7 is a flowchart illustrating a processing procedure in the learning phase of the information processing apparatus according to the present embodiment.

[FIG. 8] FIG. 8 is a flowchart illustrating a processing procedure in the analysis phase of the information processing apparatus according to the present embodiment.

[FIG. 9] FIG. 9 is a diagram (1) for describing other processing of the information processing apparatus.

[FIG. 10] FIG. 10 is a diagram (2) for describing other processing of the information processing apparatus.

[FIG. 11] FIG. 11 is a diagram illustrating an example of a hardware configuration of a computer that implements functions similar to functions of the information processing apparatus according to the embodiment.

Description of Embodiments

[0012]

Hereinafter, embodiments of an information processing program, an information processing method, and an information processing apparatus disclosed in the present application will be described in detail with reference to the drawings. Note that this invention is not limited by the embodiments.

[EMBODIMENTS]

[0013]

Processing of the information processing apparatus according to the present embodiment will be described. After processing in a learning phase is executed,

the information processing apparatus executes processing in an analysis phase. FIG. 1 is a diagram for describing the processing in the learning phase of the information processing apparatus according to the present embodiment.

[0014]

In the learning phase, the information processing apparatus executes learning of a machine learning model 50 (trains the machine learning model 50) by using a plurality of texts included in teaching data 141. The machine learning model 50 is a neural network (NN) such as pre-training of Deep Bidirectional Transformers for Language Understanding (BERT), Next Sentence Prediction, and Transformers.

[0015]

The text included in the teaching data 141 include a plurality of sentences. The plurality of sentences has a predetermined relationship with preceding and following sentences. Each sentence is a sentence set in advance based on a syllogism of an inductive method or a deductive method, or the like.

[0016]

For example, in a text 10a, a sentence "Birds lay eggs.", a sentence "Penguins are birds.", ···, and a sentence "Therefore, penguins lay eggs." in order from a first sentence. In a text 10b, a sentence "Birds are born from eggs.", a sentence "Pigeons are members of birds.", ···, and a sentence "Therefore, pigeons are born from eggs." in order from a first sentence.

[0017]

The information processing apparatus calculates sentence vectors of respective sentences included in the texts 10a and 10b and other texts. For example, the information processing apparatus calculates a sentence vector by executing morphological analysis on a sentence to divide the sentence into words and integrating the vectors of the respective words.

[0018]

A sentence vector of the sentence "Birds lay eggs." of the text 10a is "SV1-1". A sentence vector of the sentence "Penguins are birds" is "SV1-2". A sentence

vector of the sentence "Therefore, penguins lay eggs." is "SV1-3".

[0019]

A sentence vector of the sentence "Birds are born from eggs." of the text 10b is "SV2-1". A sentence vector of the sentence "Pigeons are members of birds." is "SV2-2". A sentence vector of the sentence "Therefore, pigeons lay eggs." is "SV2-3".

[0020]

The information processing apparatus repeatedly executes processing of inputting vectors to the machine learning model 50 in order from the vector of the first sentence included in the text. For example, the information processing apparatus inputs the sentence vectors to the machine learning model 50 in the order of the sentence vectors "SV1-1", "SV1-2", ···, and "SV1-3". The information processing apparatus inputs the sentence vectors to the machine learning model 50 in the order of the sentence vectors "SV2-1", "SV2-2", ···, and "SV2-3".

[0021]

By the information processing apparatus executing the processing in the learning phase described above, the machine learning model 50 that predicts a sentence vector of a second sentence next to a certain first sentence in a case where a sentence vector of the first sentence is input is generated.

[0022]

FIG. 2 is a diagram for describing the processing in the analysis phase of the information processing apparatus according to the present embodiment. In the analysis phase, the information processing apparatus calculates a sentence vector included in a text to be processed by using the trained machine learning model 50, and detects an inappropriate sentence based on a cosine similarity or the like.

[0023]

In the description of FIG. 2, a text to be processed including an input error or the like is referred to as a text 20. The text 20 includes a sentence "Birds lay eggs.", a sentence "Penguins are photographed.", ···, and a sentence "Therefore, penguins lay eggs." in order from a first sentence. The sentence "Penguins are

photographed." is a sentence including an input error of "photographed" which is a homonym of a word "birds" for a correct sentence "Penguins are birds." included in the text 10a of the teaching data 141.

[0024]

The information processing apparatus predicts a sentence vector of a next sentence of the sentence "Birds lay eggs." by calculating the sentence vector "SV1-1" of the sentence "Birds lay eggs." and inputting the calculated sentence vector "SV1-1" to the machine learning model 50. In the example illustrated in FIG. 2, "SV1-2" is predicted as the sentence vector of the next sentence of the sentence "Birds lay eggs." by the machine learning model 50.

[0025]

The information processing apparatus calculates a sentence vector "SV3" of "Penguins are photographed", which is the sentence included in the text 20 and is next to the sentence "Birds lay eggs.".

[0026]

The information processing apparatus calculates a cosine similarity between the sentence vector "SV1-2" of the next sentence predicted by the machine learning model 50 and the sentence vector "SV3" of the sentence "Penguins are photographed", which is included in the text 20 and is next to the sentence "Birds lay eggs.".

[0027]

The information processing apparatus determines that the sentence "Penguins are birds.", which is included in the text 10a and is next to the sentence "Birds lay eggs", is a correct (hereinafter, referred to as appropriate) sentence as a case where the cosine similarity is less than a threshold value. On the other hand, the information processing apparatus determines that the "Penguins are photographed.", which is included in the text 20 and is next to the sentence "Birds lay eggs.", is an inappropriate sentence including an input error or the like as a case where the cosine similarity is less than the threshold value.

[0028]

As described above, the vectors of the respective sentences of the text included in the teaching data 141 are sequentially input to the machine learning model 50, and thus, the information processing apparatus generates the machine learning model 50 that predicts the sentence vector of the second sentence next to the certain first sentence in a case where the sentence vector of the first sentence is input. The information processing apparatus inputs a sentence vector of a sentence of a text to be processed to the generated machine learning model, predicts a sentence vector of a next sentence, and detects a sentence including an input error from the text to be processed based on the predicted sentence vector. That is, a sentence including an input error or the like and having an inappropriate sentence vector may be detected from the sentences included in the text to be processed.

[0029]

Note that in the processing of FIG. 2, in a case where it is determined that the sentence "Penguins are photographed" is a sentence having an inappropriate sentence vector, the information processing apparatus may search for the sentence "Penguins are birds." having an appropriate sentence vector from a database (DB) or the like based on the sentence vector SV1-2 predicted by the machine learning model 50, and may output (hereinafter, referred to as "correct"), as a correct revision candidate, the sentence having the appropriate sentence vector to a display device.

[0030]

By using another machine learning model that has learned the order of vectors on a word-by-word basis, the information processing apparatus may calculate respective word vectors of a plurality of words "Penguins", "are", and "photographed." constituting the sentence "Penguins are photographed" from which the inappropriate sentence vector is detected, and may correct an input error or the like of the deviating word "photographed.".

[0031]

Next, a configuration example of the information processing apparatus that

executes the processing described in FIGs. 1 and 2 will be described. FIG. 3 is a functional block diagram illustrating a configuration of the information processing apparatus according to the present embodiment. As illustrated in FIG. 3, an information processing apparatus 100 includes a communication unit 110, an input unit 120, a display unit 130, a storage unit 140, and a control unit 150.

[0032]

The communication unit 110 is coupled to an external apparatus or the like in a wired or wireless manner, and transmits and receives information to and from the external apparatus or the like. The communication unit 110 is implemented by, for example, a network interface card (NIC) or the like. The communication unit 110 may be coupled to a network (not illustrated).

[0033]

The input unit 120 is an input device that inputs various types of information to the information processing apparatus 100. The input unit 120 corresponds to a keyboard, a mouse, a touch panel, or the like. For example, a user may input data of a text or the like by operating the input unit 120.

[0034]

The display unit 130 is a display device that displays information output from the control unit 150. The display unit 130 corresponds to a liquid crystal display, an organic electroluminescence (EL) display, a touch panel, or the like. For example, a sentence having an input error is displayed on the display unit 130.

[0035]

The storage unit 140 includes the machine learning model 50, the teaching data 141, and a word vector dictionary 142. For example, the storage unit 140 is implemented by a semiconductor memory device such as a random-access memory (RAM) or a flash memory, or a storage device such as a hard disk or an optical disk.

[0036]

The machine learning model 50 is an NN or the like such as BERT, Next Sentence Prediction, or Transformers described in FIG. 1.

[0037]

The teaching data 141 is the teaching data 141 described in FIG. 1. The text included in the teaching data 141 include a plurality of sentences. The plurality of sentences have a predetermined relationship with preceding and following sentences. Each sentence is a sentence set in advance based on a syllogism of an inductive method or a deductive method, or the like.

[0038]

The word vector dictionary 142 is a table that defines codes and word vectors assigned to words. FIG. 4 is a diagram illustrating an example of a data structure of the word vector dictionary. As illustrated in FIG. 4, the word vector dictionary 142 includes a code, a word, and word vectors (1) to (7). The code is a code assigned to a word. The word is the word included in a character string. The word vectors (1) to (7) are vectors assigned to the word. An N-th component of the word vector is denoted as a word vector (n) (n = 1 to 7).

[0039]

A DB 143 has various texts. A text includes a plurality of sentences, and each sentence includes a plurality of words. The DB 143 may have the texts included in the teaching data 141.

[0040]

A sentence transposition index 144 associates a sentence vector with a position pointer. The position pointer indicates a position in the DB 143 where a sentence corresponding to the sentence vector is present.

[0041]

The description returns to FIG. 3. The control unit 150 includes a pre-processing unit 151, a learning unit 152, and an analysis unit 153. For example, the control unit 150 is implemented by a central processing unit (CPU) or a micro processing unit (MPU). In addition, the control unit 150 may also be executed by, for example, an integrated circuit such as an application-specific integrated circuit (ASIC) or a field-programmable gate array (FPGA).

[0042]

The pre-processing unit 151 executes various types of pre-processing. For example, the pre-processing unit 151 acquires a sentence from the DB 143 in an unprocessed manner, and calculates a sentence vector of the sentence. The pre-processing unit 151 sets a relationship between the calculated sentence vector and the position pointer of the sentence corresponding to the sentence vector in the sentence transposition index 144.

[0043]

An example of processing of, by the pre-processing unit 151, calculating the sentence vector of the sentence will be described. FIGs. 5A and 5B are diagrams for describing the processing of calculating the sentence vector. Herein, a case where a sentence vector of a sentence "A horse does like a carrot." is calculated will be described. The pre-processing unit 151 decomposes the sentence "A horse does like a carrot." into a plurality of words by executing morphological analysis. "△ (space)" is given to each of the decomposed words. For example, Sentence 1 "A horse does like a carrot." is divided into "A△", "horse△", "does△", "like△", "a△", "carrot△", and ".△".

[0044]

The pre-processing unit 151 specifies codes corresponding to the respective words and replaces the words with the respective codes, by comparing the individual divided words with the word vector dictionary 45. For example, the individual words "A△", "horse△", "does△", "like△", "a△", "carrot△", and ".△" are replaced with "C1", "C2", "C3", "C4", "C5", "C6", and "C7", respectively.

[0045]

Description proceeds to FIG. 5B. The pre-processing unit 151 specifies the word vectors (1) to (7) assigned to the codes based on the word vector dictionary 45 and the individual codes. For example, the word vectors (1) to (7) of the code "C1" are wv1-1 to wv1-7. The word vectors (1) to (7) of the code "C2" are wv2-1 to wv2-7. The word vectors (1) to (7) of the code "C3" are wv3-1 to wv3-7.

[0046]

The word vectors (1) to (7) of the code "C4" are wv4-1 to wv4-7. The word

vectors (1) to (7) of the code "C5" are wv5-1 to wv5-7. The word vectors (1) to (7) of the code "C6" are wv6-1 to wv6-7. The word vectors (1) to (7) of the code "C7" are wv7-1 to wv7-7.

[0047]

The pre-processing unit 151 calculates the sentence vector SV1 of the sentence by integrating the word vectors for each element. For example, the pre-processing unit 151 calculates a first component "SV1-1" of the sentence vector SV1 by integrating wv1-1 to wv7-1 serving as the respective word vectors (1). The pre-processing unit 151 calculates a second component "SV1-2" of the sentence vector SV1 by integrating wv1-2 to wv7-2 serving as the respective word vectors (2). The pre-processing unit calculates a third component "SV1-3" of the sentence vector SV1 by integrating wv1-3 to wv7-3 serving as the respective word vectors (3).

[0048]

The pre-processing unit 151 calculates a fourth component "SV1-4" of the sentence vector SV1 by integrating wv1-4 to wv7-4 serving as the respective word vectors (4). The pre-processing unit 151 calculates a fifth component "SV1-5" of the sentence vector SV1 by integrating wv1-5 to wv7-5 serving as the respective word vectors (5). The pre-processing unit 151 calculates a sixth component "SV1-6" of the sentence vector SV1 by integrating wv1-6 to wv7-6 serving as the respective word vectors (6). The pre-processing unit 151 calculates a seventh component "SV1-7" of the sentence vector SV1 by integrating wv1-7 to wv7-7 serving as the respective word vectors (7).

[0049]

The above-described processing is repeatedly executed on individual sentences of the other texts included in the DB 143, and thus, the pre-processing unit 151 calculates sentence vectors of the respective sentences.

[0050]

The pre-processing unit 151 generates the sentence transposition index 144 by associating the calculated sentence vectors of the respective sentences with the

position pointers of the DB 143. Note that the pre-processing unit 151 may generate the sentence transposition index 144 having a data structure illustrated in FIG. 6. FIG. 6 is a diagram for describing processing of generating the sentence transposition index. As illustrated in FIG. 6, the pre-processing unit 151 may associate the sentence vectors, a plurality of record pointers, and a plurality of position pointers with each other, and may associate the individual record pointers and the position pointers with the respective sentences of the DB 143.

[0051]

The description returns to FIG. 3. By executing the processing in the learning phase described in FIG. 1, the learning unit 152 generates the machine learning model 50 that predicts the sentence vector of the second sentence next to the certain first sentence in a case where the sentence vector of the first sentence is input.

[0052]

For example, the learning unit 152 executes the learning of the machine learning model 50 by calculating the sentence vectors of the respective sentences included in the text of the teaching data 141 and sequentially inputting the calculated sentence vectors to the machine learning model 50. Other processing of the learning unit 152 is similar to the processing described in FIG. 1. Processing of, by the learning unit 152, calculating the sentence vectors of the sentences is similar to the processing of, by the pre-processing unit 151, calculating the sentence vectors of the sentences.

[0053]

By executing the processing in the analysis phase described in FIG. 2, the analysis unit 153 detects a sentence having an inappropriate sentence vector from the sentences included in the text to be processed.

[0054]

For example, in a case where the text 20 to be processed is received, the analysis unit 153 calculates the sentence vectors of the sentences included in the text 20. The analysis unit 153 specifies the sentences included in the text 20 based

on periods "." included in the text 20. Processing of, by the analysis unit 153, calculating the sentence vectors of the sentences is similar to the processing of, by the pre-processing unit 151, calculating the sentence vectors of the sentences. A sentence vector of an n-th sentence from the first sentence of the text 20 is denoted by "SVn" (n = 0 to M).

[0055]

The analysis unit 153 predicts a sentence vector SVn+1' of an (n+1)-th sentence from the first sentence of the text 20 by inputting the sentence vector SVn to the trained machine learning model 50. The analysis unit 153 calculates a cosine similarity between the sentence vector SVn+1' predicted by using the machine learning model 50 and a vector SVn+1 of the sentence.

[0056]

In a case where the cosine similarity between the sentence vector SVn+1' and the sentence vector SVn+1 is greater than or equal to a threshold value, the analysis unit 153 determines that the (n+1)-th sentence from the first sentence is an appropriate sentence. On the other hand, in a case where the cosine similarity between the sentence vector SVn+1' and the sentence vector SVn+1 is greater than or equal to the threshold value, the analysis unit 153 determines that the (n+1)-th sentence from the first sentence is a sentence having an inappropriate sentence vector.

[0057]

In a case where it is determined that the sentence is the sentence having the inappropriate sentence vector, the analysis unit 153 compares the sentence vector SVn+1' with the sentence transposition index 144, and specifies the position pointer of the sentence corresponding to the sentence vector SVn+1'. The analysis unit 153 searches for the sentence corresponding to the sentence vector SVn+1' from the DB 143 based on the position pointer. The analysis unit 153 displays, on the display unit 130, the sentence having the inappropriate sentence vector and the searched sentence in association with each other.

[0058]

The analysis unit 153 may compare the sentence having the inappropriate sentence vector with the searched sentence on a word-by-word basis, may detect a word with an input error from the sentence having the inappropriate sentence vector, and may display the detected word.

[0059]

Next, an example of a processing procedure of the information processing apparatus 100 according to the present embodiment will be described. FIG. 7 is a flowchart illustrating a processing procedure in the learning phase of the information processing apparatus according to the present embodiment. As illustrated in FIG. 7, the learning unit 152 of the information processing apparatus 100 selects an unselected text from the teaching data 141 (step S101).

[0060]

The learning unit 152 calculates sentence vectors of respective sentences included in the selected text, and generates a sentence transposition index in which the sentence vectors, the records in the DB, and the positions of the sentences are associated with each other (step S102). The learning unit 152 executes learning by sequentially inputting the sentence vectors to the machine learning model 50 from the sentence vector of the first sentence included in the selected text (step S103).

[0061]

In a case where learning is continued (Yes in step S104), the learning unit 152 proceeds to step S101. In a case where learning is not continued (No in step S104), the learning unit 152 ends the processing in the learning phase.

[0062]

FIG. 8 is a flowchart illustrating a processing procedure in the analysis phase of the information processing apparatus according to the present embodiment. As illustrated in FIG. 8, the analysis unit 153 of the information processing apparatus 100 receives an input of a text to be processed (step S201).

[0063]

The analysis unit 153 calculates individual sentence vectors of respective

sentences included in the input text (step S202). The analysis unit 153 sets n to an initial value (step S203).

[0064]

The analysis unit 153 predicts the sentence vector SVn+1' of the (n+1)-th sentence by inputting the sentence vector SVn of the n-th sentence among the plurality of sentences included in the text to the machine learning model 50 (step S204).

[0065]

The analysis unit 153 calculates the cosine similarity between the sentence vector SVn+1 of the (n+1)-th sentence and the predicted sentence vector SVn+1' of the sentence among the plurality of sentences included in the text (step S205).

[0066]

In a case where the cosine similarity is greater than or equal to the threshold value (Yes in step S206), the analysis unit 153 proceeds to step S210.

[0067]

On the other hand, in a case where the cosine similarity is not greater than or equal to the threshold value (No in step S206), the analysis unit 153 detects the (n+1)-th sentence as the sentence having the inappropriate sentence vector (step S207). The analysis unit 153 detects the sentence corresponding to the sentence vector SVn+1' from the DB 143 based on the predicted sentence vector SVn+1' and the sentence transposition index 144 (step S208).

[0068]

The analysis unit 153 displays, on the display unit 130, the sentence having the inappropriate sentence vector and the sentence detected from the DB 143 (step S209).

[0069]

Processing in step S210 and subsequent steps will be described. In a case where n is greater than or equal to L (Yes in step S210), the analysis unit 153 ends the processing. L is the number of sentences included in the text to be processed. In a case where n is not greater than or equal to L (No in step S210), the analysis

unit 153 updates n by a value obtained by adding 1 to n (step S211), and proceeds to step S204.

[0070]

Next, effects of the information processing apparatus 100 according to the present embodiment are described. The information processing apparatus 100 generates the machine learning model 50 that predicts the sentence vector of the second sentence next to the certain first sentence in a case where the sentence vector of the first sentence is input by sequentially inputting the vectors of the respective sentences of the text included in the teaching data 141 to the machine learning model 50. The information processing apparatus 100 predicts the sentence vector of the next sentence by inputting the sentence vectors of the sentences of the text to be processed to the generated machine learning model 50, and detects the sentence having the inappropriate sentence vector from the text to be processed based on the predicted sentence vector. In addition, the word with the input error or the like may be corrected from the inappropriate sentence.

[0071]

The information processing apparatus 100 detects the sentence having the inappropriate sentence vector based on the cosine similarity between the sentence vector of the next sentence predicted by the machine learning model 50 and the sentence vector of the next sentence of the sentence included in the text to be processed, and corrects the input error or the like. Accordingly, calculation cost is suppressed, the sentence having the inappropriate sentence vector is detected, and thus, the input error or the like may be corrected.

[0072]

The information processing apparatus 100 sequentially inputs, to the machine learning model, the vectors of the plurality of sentences, the arrangement order of which is determined based on the inductive method or the deductive method, and trains the machine learning model. Accordingly, a next sentence of a target sentence may be predicted based on the inductive method or the deductive method.

[0073]

In a case where it is determined that the sentence is a sentence to be revised, the information processing apparatus 100 searches for a corrected sentence based on the vector predicted by the machine learning model 50. Accordingly, the revised sentence may be notified.

[0074]

Note that the processing content of the information processing apparatus 100 described above is an example, and the information processing apparatus 100 may execute other processing. The other processing of the information processing apparatus 100 will be described below.

[0075]

FIGs. 9 and 10 are diagrams for describing the other processing of the information processing apparatus. Although the above-described information processing apparatus 100 causes the machine learning model 50 to learn the order of sentence vectors based on the syllogism, the information processing apparatus may cause the machine learning model to learn the order of vectors of a protein primary structure, which is a sequence of a protein and is constituted by a plurality of amino acid sequences corresponding to words, instead of the sentence vectors. In the following description, a continuous amino acid sequence of a protein is referred to as a "basic structure", and a protein primary structure is referred to as a "primary structure".

[0076]

FIG. 9 is described. In the learning phase, the information processing apparatus 100 executes learning of the machine learning model 50 by using sequences 20a and 20b of a plurality of proteins included in teaching data 241.

[0077]

For example, the sequence 20a includes primary structures "α primary structure", "β primary structure", ···, and "γ primary structure". The sequence 20b includes primary structures "Δ primary structure", "ε primary structure", ···, and "ζ primary structure".

[0078]

The information processing apparatus 100 specifies vectors of respective primary structures by using a vector dictionary of basic structures of proteins in which basic structures and vectors are associated with each other. For example, a vector of the primary structure "α primary structure" constituted by a plurality of basic structures is "V20-1", a vector of the primary structure "β primary structure" is "V20-2", and a vector of the primary structure "γ primary structure" is "V20-3". A vector of the primary structure "Δ primary structure" is "V21-1", a vector of the primary structure "ε primary structure" is "V21-2", and a vector of the primary structure "ζ primary structure" is "V21-3". The vector of each primary structure is calculated based on the vectors of the respective basic structures of the plurality of basic structures constituting that primary structure.

[0079]

The information processing apparatus 100 repeatedly executes processing of inputting the vectors to the machine learning model 50 in order from a vector of a first primary structure included in the sequence of the protein. For example, the information processing apparatus inputs the vectors to the machine learning model 50 in the order of the vectors "V20-1", "V20-2", ···, and "V20-3". The information processing apparatus inputs the vectors to the machine learning model 50 in the order of the vectors "V21-1", "V21-2", ···, and "V21-3".

[0080]

By the information processing apparatus 100 executing the processing in the learning phase described above, the machine learning model 50 that predicts a vector of a primary structure next to a certain primary structure in a case where a vector of the primary structure is input is generated.

[0081]

FIG. 10 is described. In the analysis phase, a sequence of a protein to be processed is a sequence 25. Primary structures "α primary structure", "η primary structure", ···, and "γ primary structure" are included in the sequence 25 in order from a first primary structure.

[0082]

The information processing apparatus 100 predicts a vector of a primary structure next to the primary structure "α primary structure" by calculating the vector "V20-1" of the primary structure "α primary structure" and inputting the calculated vector "V20-1" to the machine learning model 50. In the example illustrated in FIG. 10, "V20-2" is predicted by the machine learning model 50, as the vector of the primary structure next to the primary structure "α primary structure".

[0083]

The information processing apparatus 100 calculates a vector "V22" of the "η primary structure", which is a primary structure included in the sequence 25 and is next to the primary structure "α primary structure".

[0084]

The information processing apparatus 100 calculates a cosine similarity between the vector "V20-2" of the next primary structure predicted by the machine learning model 50 and the vector "V22" of the "η primary structure" next to the basic structure "α primary structure", which is the primary structure included in the sequence 25.

[0085]

In a case where the cosine similarity is greater than or equal to a threshold value, the information processing apparatus determines that the "η primary structure", which is the primary structure included in the sequence 25 and is next to the primary structure "α primary structure", is a correct primary structure. On the other hand, in a case where the cosine similarity is less than the threshold value, the information processing apparatus determines that the "η primary structure", which is the primary structure included in the sequence 25 and is next to the primary structure "α primary structure", is an inappropriate primary structure, and corrects a mutation or the like of the basic structure included in the primary structure "η primary structure".

[0086]

By the information processing apparatus 100 executing the processing illustrated in FIGs. 9 and 10, a primary structure having an inappropriate primary structure vector may be detected from the plurality of primary structures included in the sequence of the protein, and a basic structure having a mutation or the like may be corrected. Accordingly, a protein primary structure having a mutation or the like (SNPs is a representative example) occurring in a receptor constituted by a plurality of protein primary structures may be detected. Further, by performing machine learning on a large number of protein primary structures constituting the receptor and a single or a plurality of protein primary structures bound to the receptor in order of binding, a vector of a protein primary structure of a ligand bound to the receptor may be predicted. Accordingly, improvement of a ligand having a vector of a new protein primary structure, which is similar to a ligand already commercialized as a biopharmaceutical, has an excellent medicinal effect, and in which side reactions are suppressed may be supported.

[0087]

Next, an example of a hardware configuration of a computer that implements functions similar to the functions of the information processing apparatus 100 illustrated in the above embodiment will be described. FIG. 11 is a diagram illustrating an example of the hardware configuration of the computer that implements the functions similar to the functions of the information processing apparatus according to the embodiment.

[0088]

As illustrated in FIG. 11, a computer 300 includes a CPU 301 that executes various types of arithmetic processing, an input device 302 that receives input of data from a user, and a display 303. The computer 300 also includes a communication device 304 that exchanges data with an external apparatus or the like via a wired or wireless network, and an interface device 305. The computer 300 also includes a RAM 306 that temporarily stores various types of information and a hard disk device 307. Each of the individual devices 301 to 307 is coupled to a bus 308.

[0089]

The hard disk device 307 includes a pre-processing program 307a, a learning program 307b, and an analysis program 307c. In addition, the CPU 301 reads the individual programs 307a to 307c and loads the programs into the RAM 306.

[0090]

The pre-processing program 307a functions as a pre-processing process 306a. The learning program 307b functions as a learning process 306b. The analysis program 307c functions as an analysis process 306c.

[0091]

Processing of the pre-processing process 306a corresponds to the processing of the pre-processing unit 151. Processing of the learning process 306b corresponds to the processing of the learning unit 152. Processing of the analysis process 306c corresponds to processing of the analysis unit 153.

[0092]

Note that each of the programs 307a to 307c may not be stored in the hard disk device 307 from the beginning. For example, each program may be stored in a "portable physical medium" such as a flexible disk (FD), a CD-ROM, a DVD, a magneto-optical disk, an IC card, or the like inserted in the computer 300. The computer 300 may read and execute each of the programs 307a to 307c.

Reference Signs List

CLAIMS

1. An information processing program causing a computer to execute a process comprising:

calculating individual vectors of a plurality of continuous sentences that have a relationship with preceding and following sentences;

generating a machine learning model that predicts a sentence vector of a sentence input next to a certain sentence when a vector of the certain sentence is input to the machine learning model, by sequentially inputting the vectors of the plurality of sentences to the machine learning model and training the machine learning model;

calculating a vector of a first sentence and a vector of a second sentence next to the first sentence; and

calculating a vector of a sentence predicted to be next to the first sentence by inputting the vector of the first sentence to the machine learning model, and determining whether or not the vector of the second sentence is appropriate.

2. The information processing program according to claim 1, wherein the determining is determining whether or not the vector of the second sentence is appropriate based on a cosine similarity between the vector predicted by inputting the vector of the first sentence to the machine learning model and the vector of the second sentence.

3. The information processing program according to claim 2, wherein the plurality of continuous sentences are a plurality of sentences of which an arrangement order is determined based on an inductive method or a deductive method, and the generating of the machine learning model is sequentially inputting the vectors of the plurality of sentences of which the arrangement order is determined based on the inductive method or the deductive method to the machine learning model and training the machine learning model.

4. The information processing program according to claim 1, wherein the computer is caused to further execute a process of calculating the vector of the sentence predicted to be next to the first sentence by inputting the vector of the first sentence to the machine learning model, and recommending an appropriate sentence based on the calculated vector of the sentence predicted to be next to the first sentence in order to search for a sentence similar to the calculated vector and present the searched sentence as a candidate for the appropriate sentence, in a case where it is determined that the vector of the second sentence is inappropriate.

5. An information processing method implemented by a computer, the information processing method comprising:

calculating individual vectors of a plurality of continuous sentences that have a relationship with preceding and following sentences;

generating a machine learning model that predicts a sentence vector of a sentence input next to a certain sentence when a vector of the certain sentence is input to the machine learning model, by sequentially inputting the vectors of the plurality of sentences to the machine learning model and training the machine learning model;

calculating a vector of a first sentence and a vector of a second sentence next to the first sentence; and

calculating a vector of a sentence predicted to be next to the first sentence by inputting the vector of the first sentence to the machine learning model, and determining whether or not the vector of the second sentence is appropriate.

6. The information processing method according to claim 5, wherein the determining is determining whether or not the vector of the second sentence is appropriate based on a cosine similarity between the vector predicted by inputting the vector of the first sentence to the machine learning model and the vector of the second sentence.

7. The information processing method according to claim 6, wherein the plurality of continuous sentences are a plurality of sentences of which an arrangement order is determined based on an inductive method or a deductive method, and the generating of the machine learning model is sequentially inputting the vectors of the plurality of sentences of which the arrangement order is determined based on the inductive method or the deductive method to the machine learning model and training the machine learning model.

8. The information processing method according to claim 5, wherein the computer is caused to further execute a process of calculating the vector of the sentence predicted to be next to the first sentence by inputting the vector of the first sentence to the machine learning model, and recommending an appropriate sentence based on the calculated vector of the sentence predicted to be next to the first sentence in order to search for a sentence similar to the calculated vector and present the searched sentence as a candidate for the appropriate sentence, in a case where it is determined that the vector of the second sentence is inappropriate.

9. An information processing apparatus comprising:

a control unit configured to execute a process including:

calculating individual vectors of a plurality of continuous sentences that have a relationship with preceding and following sentences;

generating a machine learning model that predicts a sentence vector of a sentence input next to a certain sentence when a vector of the certain sentence is input to the machine learning model, by sequentially inputting the vectors of the plurality of sentences to the machine learning model and training the machine learning model;

calculating a vector of a first sentence and a vector of a second sentence next to the first sentence; and

calculating a vector of a sentence predicted to be next to the first sentence by inputting the vector of the first sentence to the machine learning model, and determining whether or not the vector of the second sentence is appropriate.

10. The information processing apparatus according to claim 9, wherein the determining is determining whether or not the vector of the second sentence is appropriate based on a cosine similarity between the vector predicted by inputting the vector of the first sentence to the machine learning model and the vector of the second sentence.

11. The information processing apparatus according to claim 10, wherein the plurality of continuous sentences are a plurality of sentences of which an arrangement order is determined based on an inductive method or a deductive method, and the generating of the machine learning model is sequentially inputting the vectors of the plurality of sentences of which the arrangement order is determined based on the inductive method or the deductive method to the machine learning model and training the machine learning model.

12. The information processing apparatus according to claim 9, wherein the computer is caused to further execute a process of calculating the vector of the sentence predicted to be next to the first sentence by inputting the vector of the first sentence to the machine learning model, and recommending an appropriate sentence based on the calculated vector of the sentence predicted to be next to the first sentence in order to search for a sentence similar to the calculated vector and present the searched sentence as a candidate for the appropriate sentence, in a case where it is determined that the vector of the second sentence is inappropriate.

FIG. 1

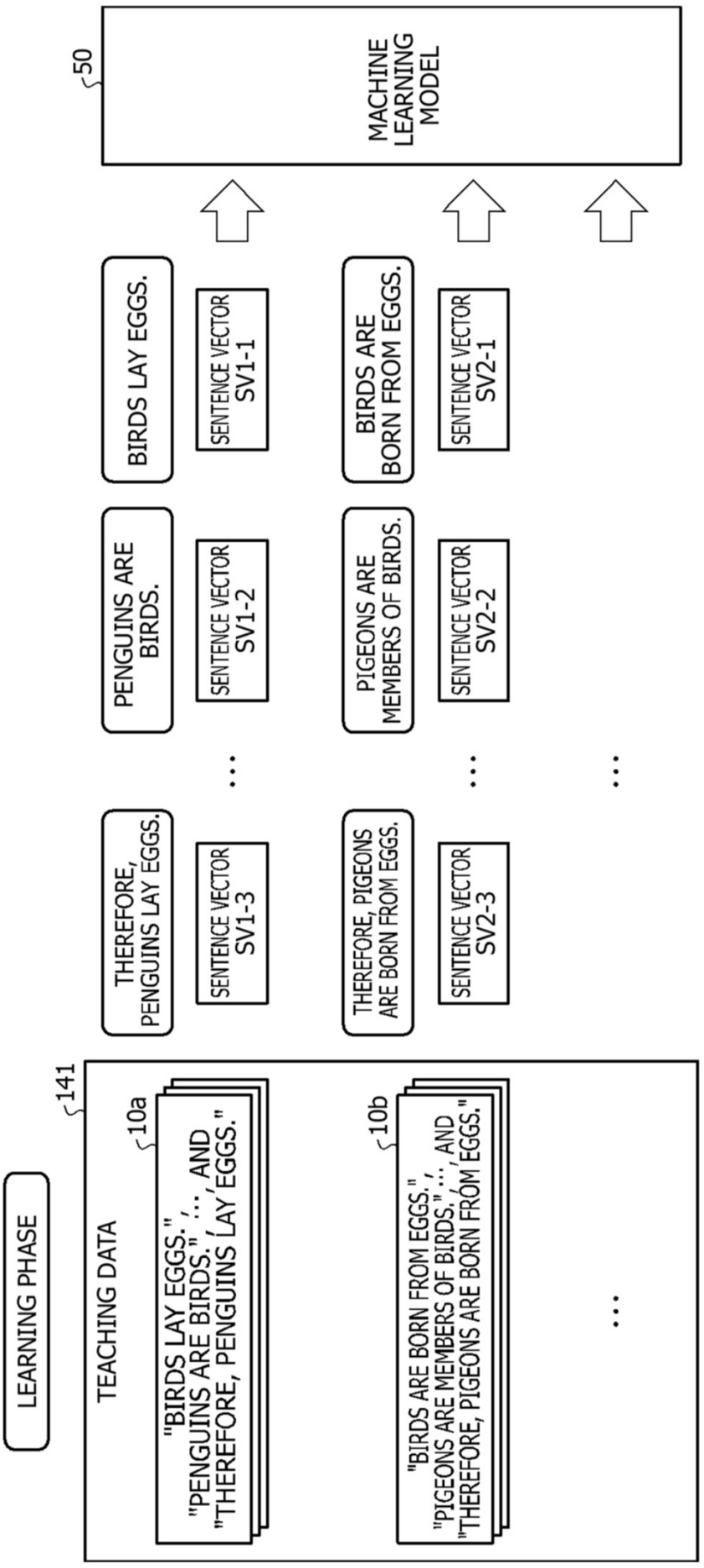

LEARNING PHASE
141
TEACHING DATA
10a
"BIRDS LAY EGGS.", "PENGUINS ARE BIRDS.", ..., AND "THEREFORE, PENGUINS LAY EGGS."
10b
"BIRDS ARE BORN FROM EGGS.", "PIGEONS ARE MEMBERS OF BIRDS.", ..., AND "THEREFORE, PIGEONS ARE BORN FROM EGGS."
THEREFORE, PENGUINS LAY EGGS.
PENGUINS ARE BIRDS.
BIRDS LAY EGGS.
THEREFORE, PIGEONS ARE BORN FROM EGGS.
PIGEONS ARE MEMBERS OF BIRDS.
BIRDS ARE BORN FROM EGGS.
SENTENCE VECTOR SV1-3
SENTENCE VECTOR SV1-2
SENTENCE VECTOR SV1-1
SENTENCE VECTOR SV2-3
SENTENCE VECTOR SV2-2
SENTENCE VECTOR SV2-1
50
MACHINE LEARNING MODEL

FIG. 2

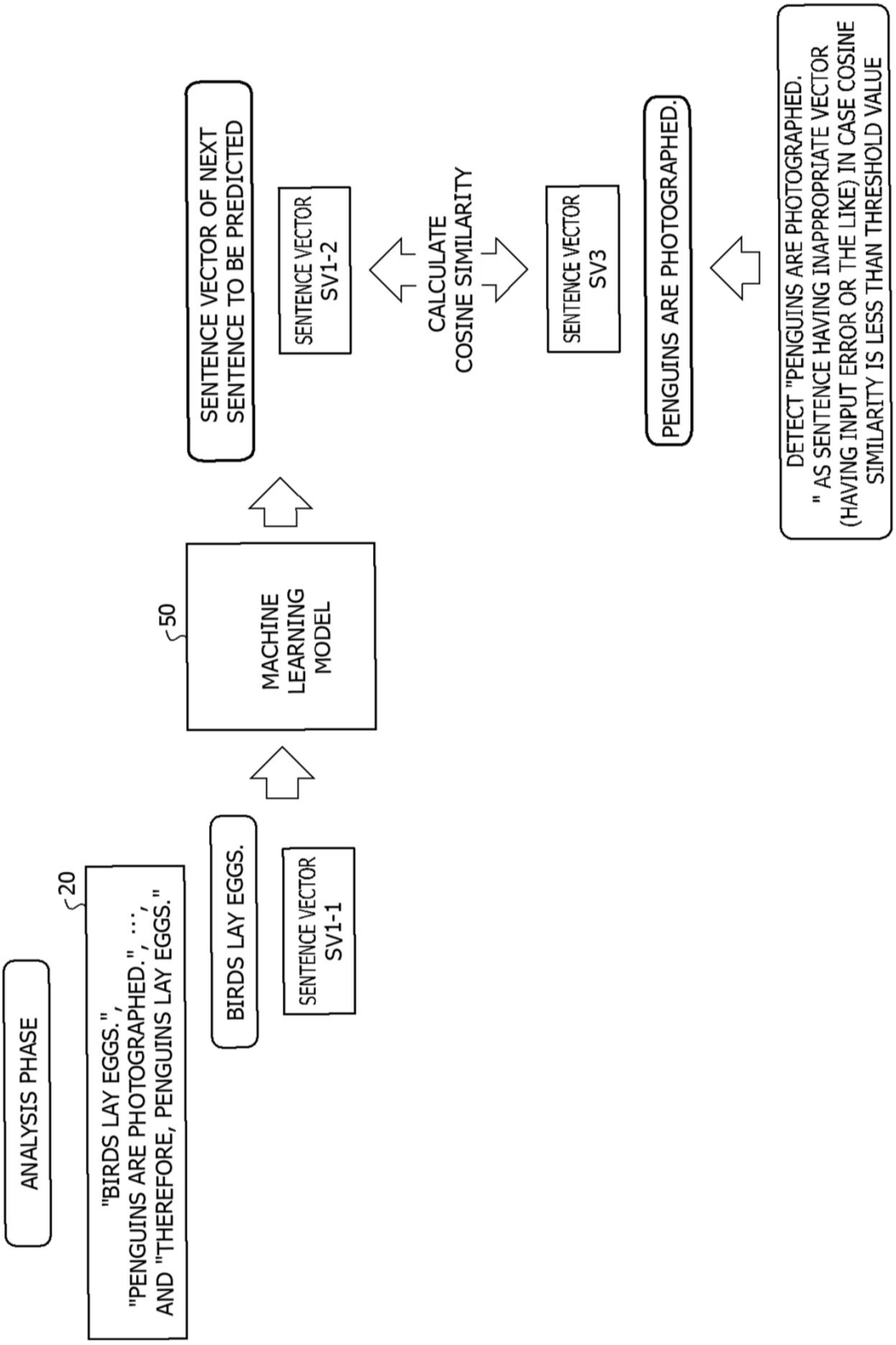

ANALYSIS PHASE
20
"BIRDS LAY EGGS.",
"PENGUINS ARE PHOTOGRAPHED.", ...,
AND "THEREFORE, PENGUINS LAY EGGS."
BIRDS LAY EGGS.
SENTENCE VECTOR SV1-1
50
MACHINE LEARNING MODEL
SENTENCE VECTOR OF NEXT SENTENCE TO BE PREDICTED
SENTENCE VECTOR SV1-2
CALCULATE COSINE SIMILARITY
SENTENCE VECTOR SV3
PENGUINS ARE PHOTOGRAPHED.
DETECT "PENGUINS ARE PHOTOGRAPHED. " AS SENTENCE HAVING INAPPROPRIATE VECTOR (HAVING INPUT ERROR OR THE LIKE) IN CASE COSINE SIMILARITY IS LESS THAN THRESHOLD VALUE

FIG. 3

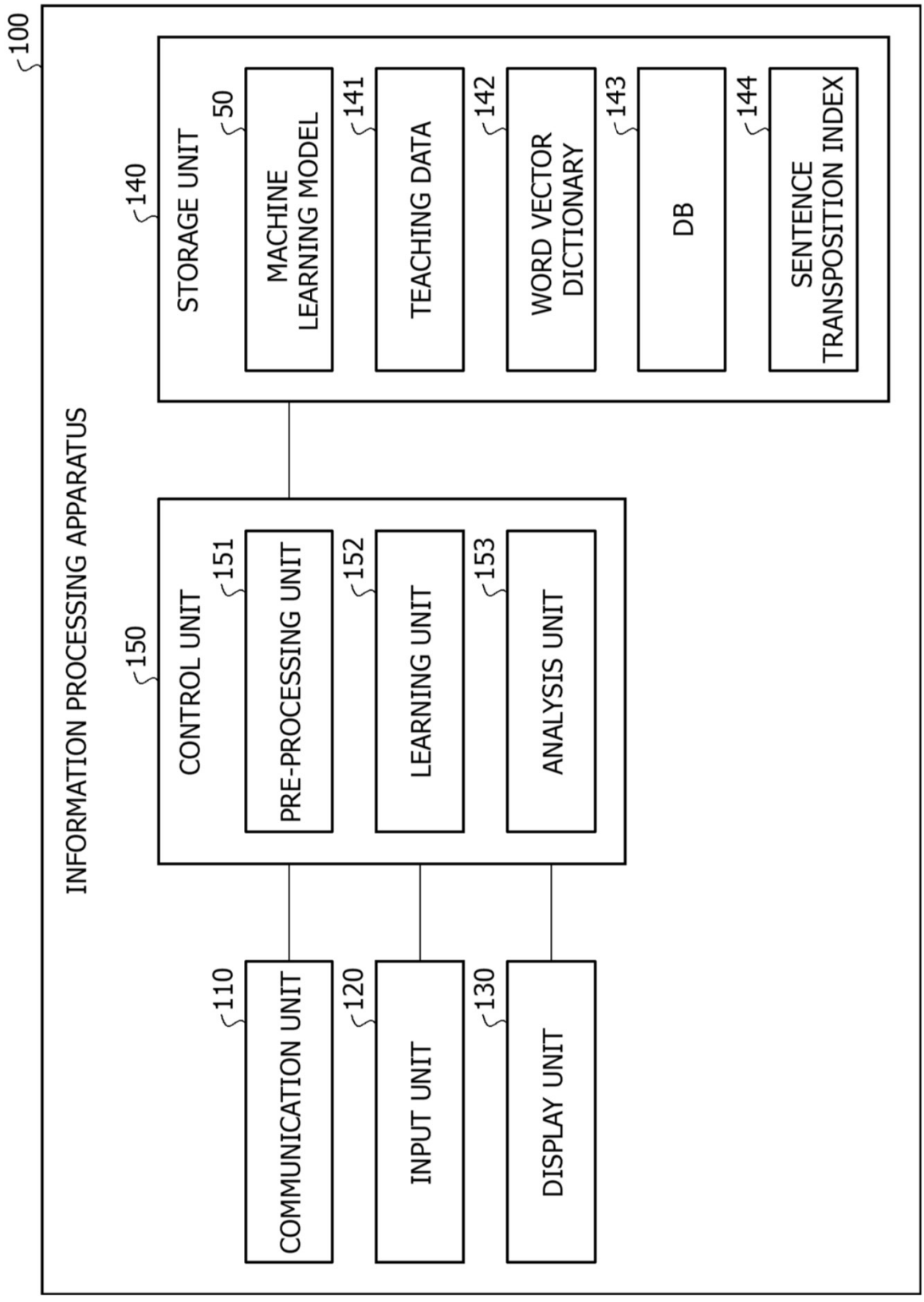

100
INFORMATION PROCESSING APPARATUS
140
STORAGE UNIT
50
MACHINE LEARNING MODEL
141
TEACHING DATA
142
WORD VECTOR DICTIONARY
143
DB
144
SENTENCE TRANSPOSITION INDEX
150
CONTROL UNIT
151
PRE-PROCESSING UNIT
152
LEARNING UNIT
153
ANALYSIS UNIT
110
COMMUNICATION UNIT
120
INPUT UNIT
130
DISPLAY UNIT

FIG. 4

142

CODE	WORD	WORD VECTOR (1)	WORD VECTOR (2)	WORD VECTOR (3)	WORD VECTOR (4)	WORD VECTOR (5)	WORD VECTOR (6)	WORD VECTOR (7)
C1	A	wv1-1	wv1-2	wv1-3	wv1-4	wv1-5	wv1-6	wv1-7
C2	HORSE	wv2-1	wv2-2	wv2-3	wv2-4	wv2-5	wv2-6	wv2-7
C3	DOES	wv3-1	wv3-2	wv3-3	wv3-4	wv3-5	wv3-6	wv3-7
C4	LIKE	wv4-1	wv4-2	wv4-3	wv4-4	wv4-5	wv4-6	wv4-7
C5	A	wv5-1	wv5-2	wv5-3	wv5-4	wv5-5	wv5-6	wv5-7
C6	CARROT	wv6-1	wv6-2	wv6-3	wv6-4	wv6-5	wv6-6	wv6-7
C7	.	wv7-1	wv7-2	wv7-3	wv7-4	wv7-5	wv7-6	wv7-7
C8	OF	wv8-1	wv8-2	wv8-3	wv8-4	wv8-5	wv8-6	wv8-7
C9	FAVORITE	wv9-1	wv9-2	wv9-3	wv9-4	wv9-5	wv9-6	wv9-7
C10	TODAY	wv10-1	wv10-2	wv10-3	wv10-4	wv10-5	wv10-6	wv10-7
C11	WEATHER	wv11-1	wv11-2	wv11-3	wv11-4	wv11-5	wv11-6	wv11-7
C12	SUNNY	wv12-1	wv12-2	wv12-3	wv12-4	wv12-5	wv12-6	wv12-7
...	...	...	...	...	...	...	...	...

FIG. 5A

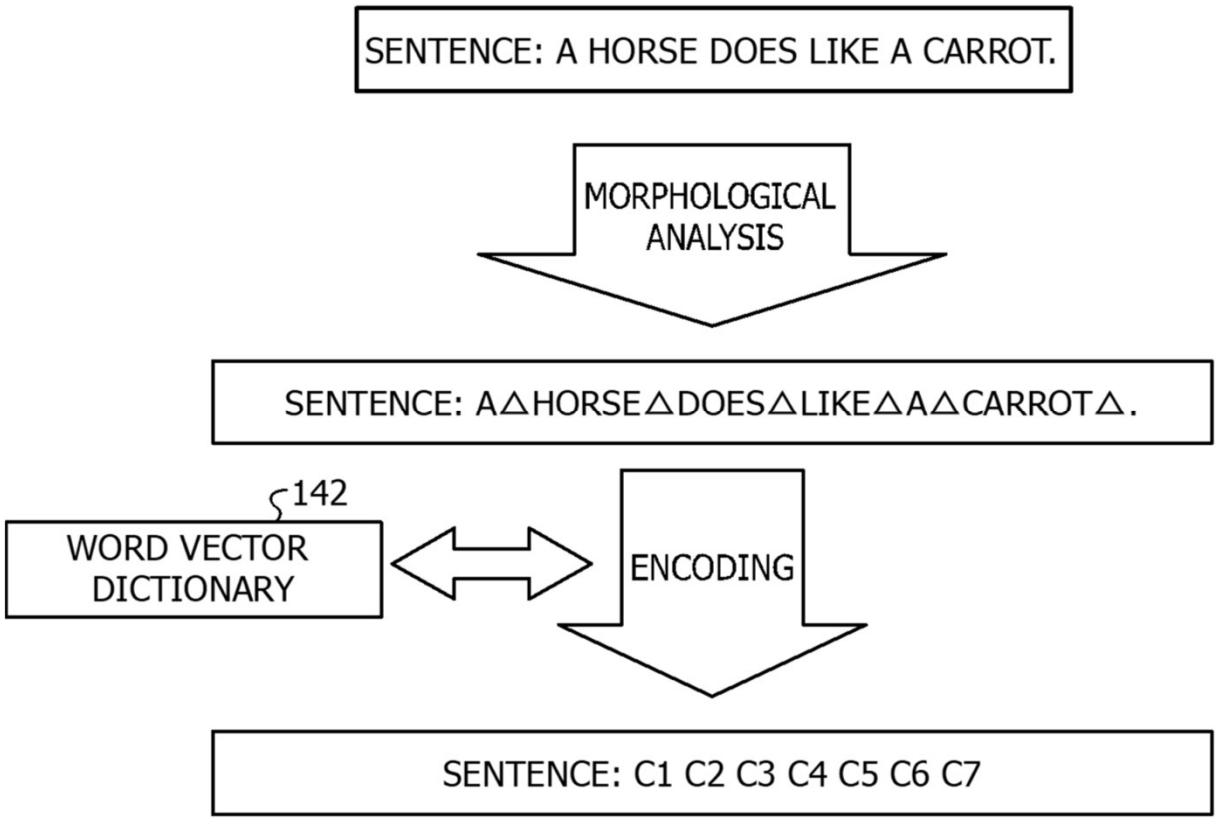

SENTENCE: A HORSE DOES LIKE A CARROT.
MORPHOLOGICAL ANALYSIS
SENTENCE: A△HORSE△DOES△LIKE△A△CARROT△.
142
WORD VECTOR DICTIONARY
ENCODING
SENTENCE: C1 C2 C3 C4 C5 C6 C7

FIG. 5B

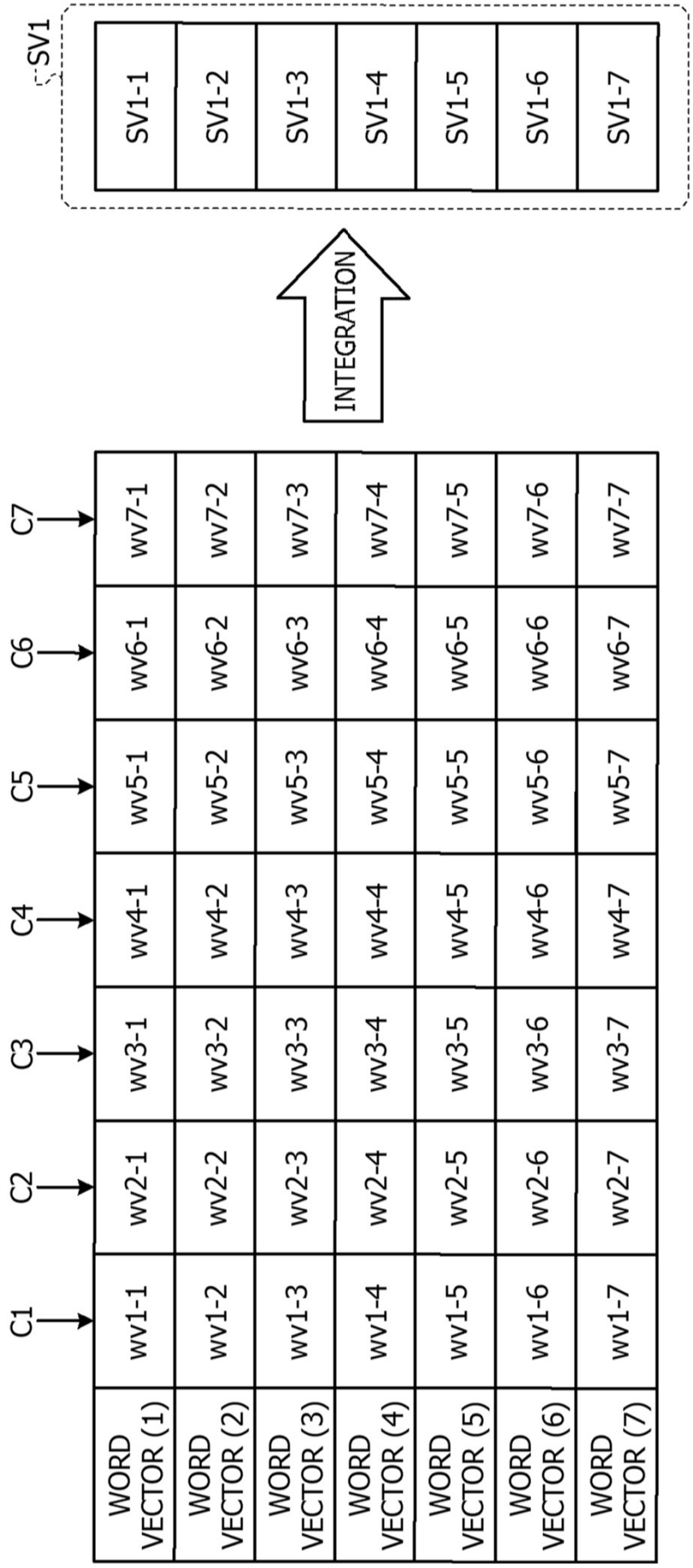

SV1
SV1-1
SV1-2
SV1-3
SV1-4
SV1-5
SV1-6
SV1-7
INTEGRATION
C1
C2
C3
C4
C5
C6
C7
WORD VECTOR (1)
WORD VECTOR (2)
WORD VECTOR (3)
WORD VECTOR (4)
WORD VECTOR (5)
WORD VECTOR (6)
WORD VECTOR (7)
wv1-1 wv2-1 wv3-1 wv4-1 wv5-1 wv6-1 wv7-1
wv1-2 wv2-2 wv3-2 wv4-2 wv5-2 wv6-2 wv7-2
wv1-3 wv2-3 wv3-3 wv4-3 wv5-3 wv6-3 wv7-3
wv1-4 wv2-4 wv3-4 wv4-4 wv5-4 wv6-4 wv7-4
wv1-5 wv2-5 wv3-5 wv4-5 wv5-5 wv6-5 wv7-5
wv1-6 wv2-6 wv3-6 wv4-6 wv5-6 wv6-6 wv7-6
wv1-7 wv2-7 wv3-7 wv4-7 wv5-7 wv6-7 wv7-7

FIG. 6

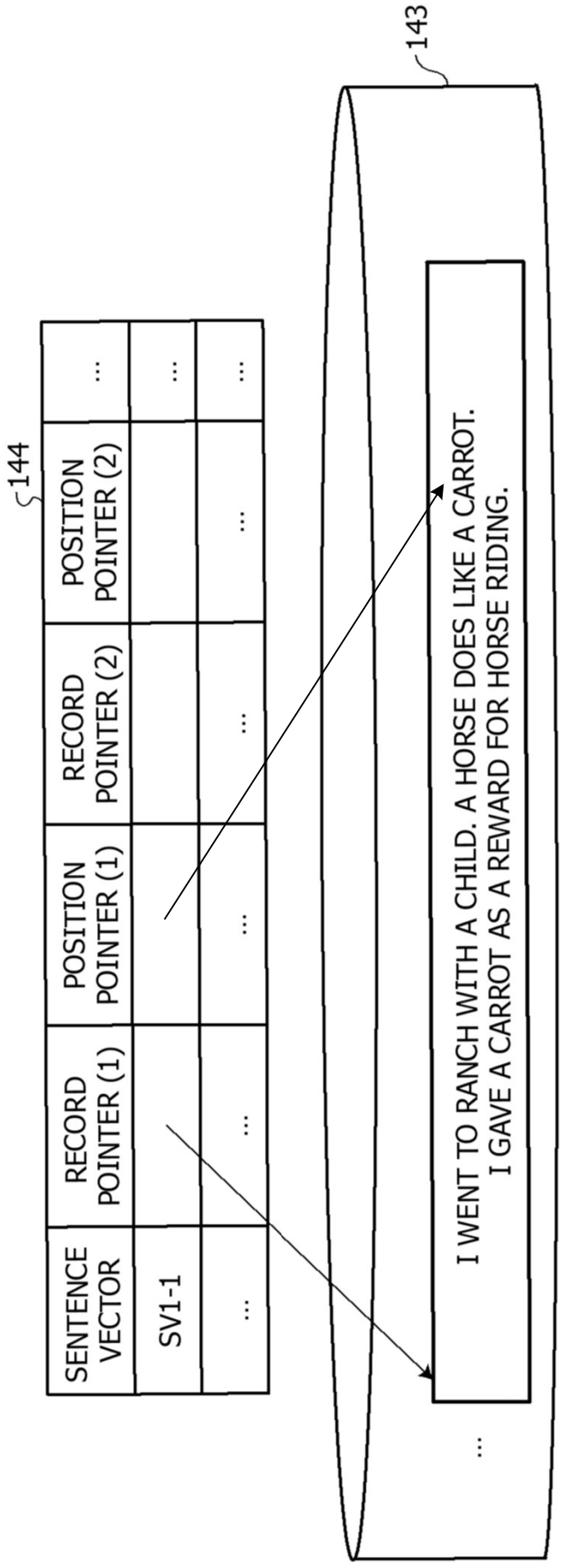

143
144
SENTENCE VECTOR
RECORD POINTER (1)
POSITION POINTER (1)
RECORD POINTER (2)
POSITION POINTER (2)
SV1-1
I WENT TO RANCH WITH A CHILD. A HORSE DOES LIKE A CARROT.
I GAVE A CARROT AS A REWARD FOR HORSE RIDING.

FIG. 7

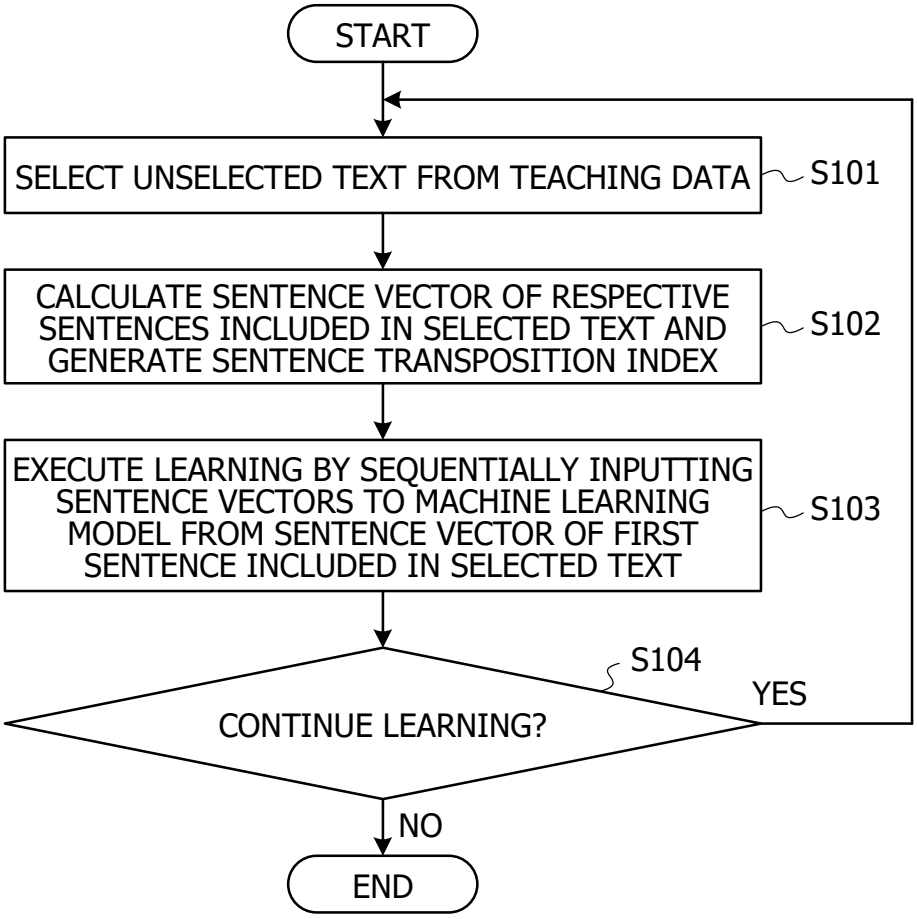

START
SELECT UNSELECTED TEXT FROM TEACHING DATA
S101
CALCULATE SENTENCE VECTOR OF RESPECTIVE SENTENCES INCLUDED IN SELECTED TEXT AND GENERATE SENTENCE TRANSPOSITION INDEX
S102
EXECUTE LEARNING BY SEQUENTIALLY INPUTTING SENTENCE VECTORS TO MACHINE LEARNING MODEL FROM SENTENCE VECTOR OF FIRST SENTENCE INCLUDED IN SELECTED TEXT
S103
S104
CONTINUE LEARNING?
YES
NO
END

FIG. 8

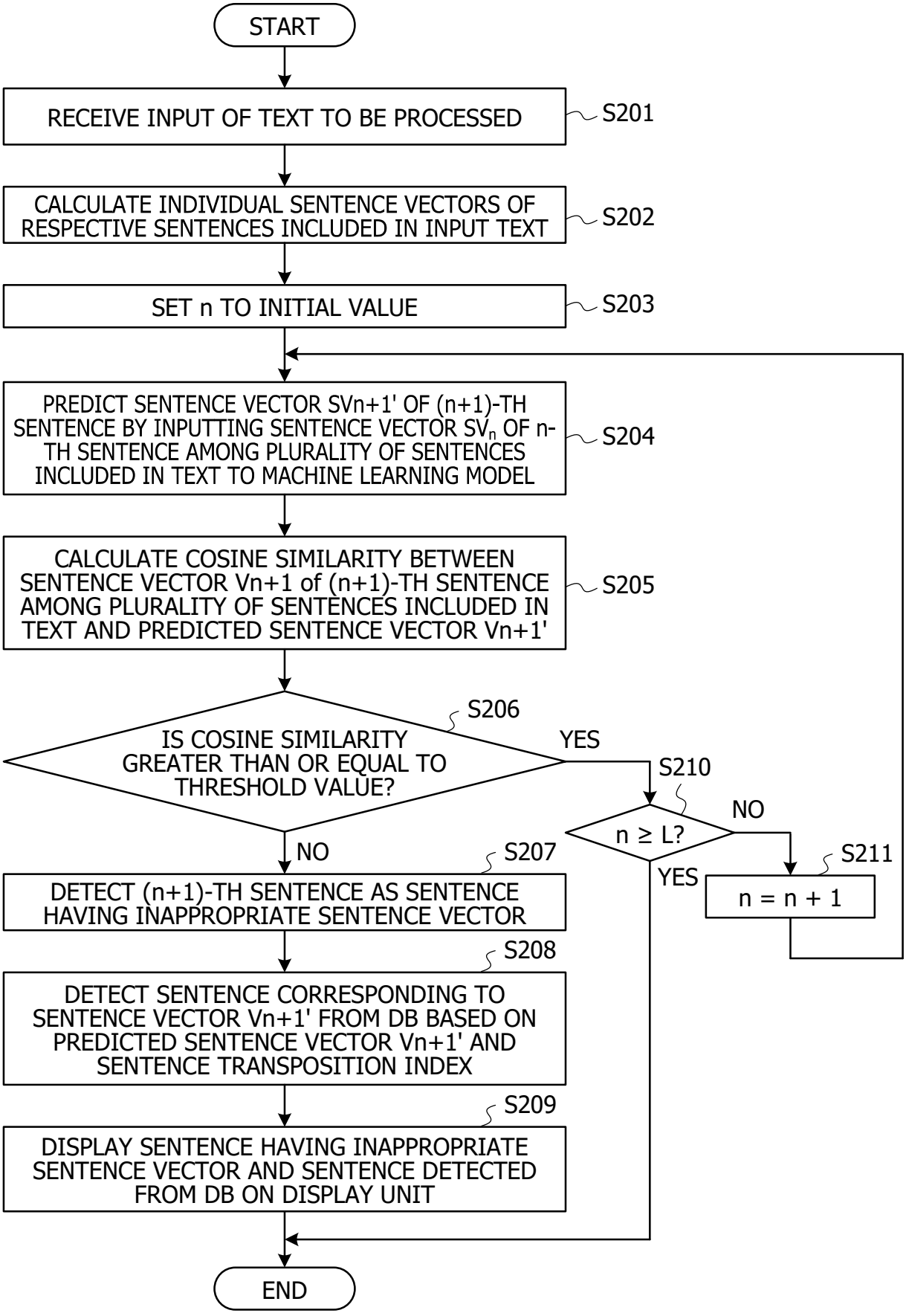

START
RECEIVE INPUT OF TEXT TO BE PROCESSED
S201
CALCULATE INDIVIDUAL SENTENCE VECTORS OF RESPECTIVE SENTENCES INCLUDED IN INPUT TEXT
S202
SET n TO INITIAL VALUE
S203
PREDICT SENTENCE VECTOR SVn+1' OF (n+1)-TH SENTENCE BY INPUTTING SENTENCE VECTOR SVn OF n-TH SENTENCE AMONG PLURALITY OF SENTENCES INCLUDED IN TEXT TO MACHINE LEARNING MODEL
S204
CALCULATE COSINE SIMILARITY BETWEEN SENTENCE VECTOR Vn+1 of (n+1)-TH SENTENCE AMONG PLURALITY OF SENTENCES INCLUDED IN TEXT AND PREDICTED SENTENCE VECTOR Vn+1'
S205
S206
IS COSINE SIMILARITY GREATER THAN OR EQUAL TO THRESHOLD VALUE?
YES
NO
S210
n ≥ L?
NO
YES
S211
n = n + 1
S207
DETECT (n+1)-TH SENTENCE AS SENTENCE HAVING INAPPROPRIATE SENTENCE VECTOR
S208
DETECT SENTENCE CORRESPONDING TO SENTENCE VECTOR Vn+1' FROM DB BASED ON PREDICTED SENTENCE VECTOR Vn+1' AND SENTENCE TRANSPOSITION INDEX
S209
DISPLAY SENTENCE HAVING INAPPROPRIATE SENTENCE VECTOR AND SENTENCE DETECTED FROM DB ON DISPLAY UNIT
END

FIG. 9

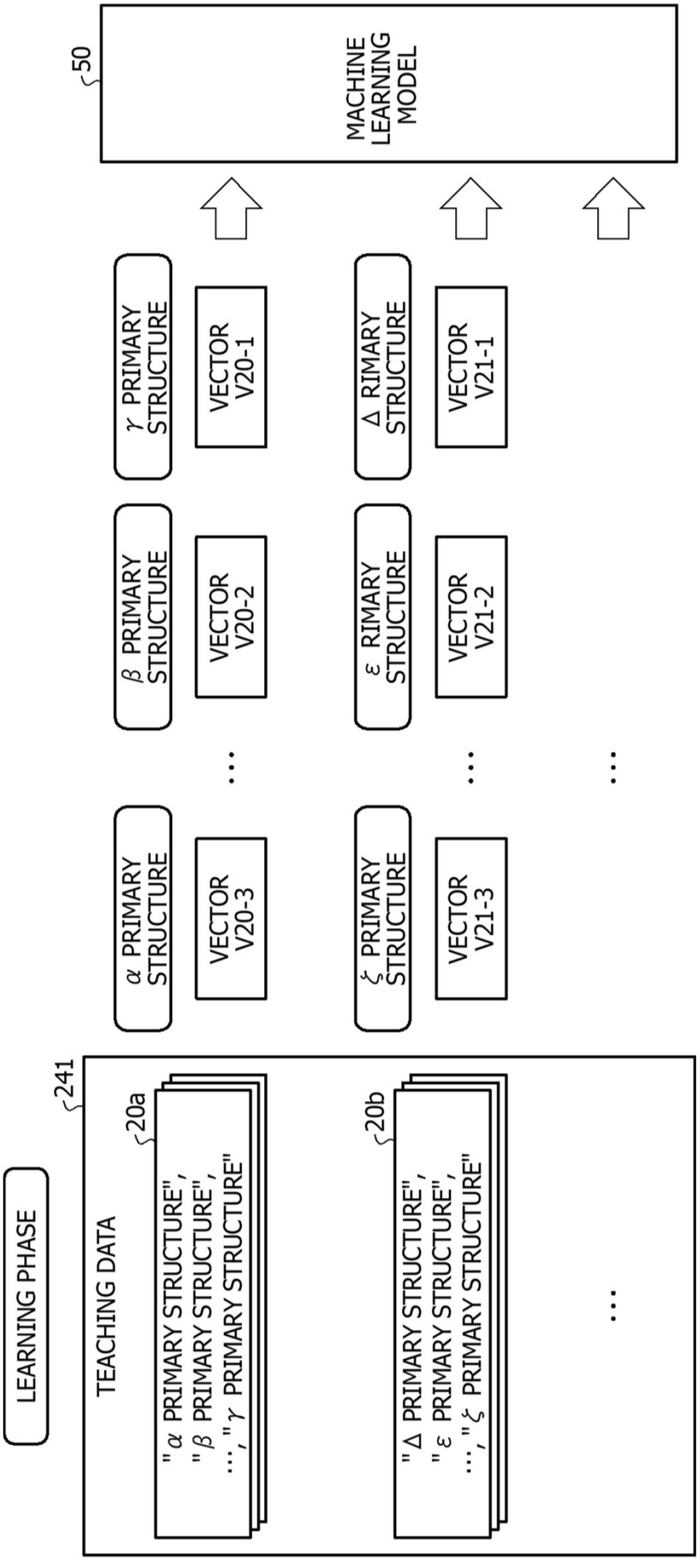

LEARNING PHASE
TEACHING DATA
241
20a
"α PRIMARY STRUCTURE",
"β PRIMARY STRUCTURE",
…,"γ PRIMARY STRUCTURE"
20b
"Δ PRIMARY STRUCTURE",
"ε PRIMARY STRUCTURE",
…,"ζ PRIMARY STRUCTURE"
α PRIMARY STRUCTURE
VECTOR V20-3
β PRIMARY STRUCTURE
VECTOR V20-2
γ PRIMARY STRUCTURE
VECTOR V20-1
ζ PRIMARY STRUCTURE
VECTOR V21-3
ε RIMARY STRUCTURE
VECTOR V21-2
Δ RIMARY STRUCTURE
VECTOR V21-1
50
MACHINE LEARNING MODEL

FIG. 10

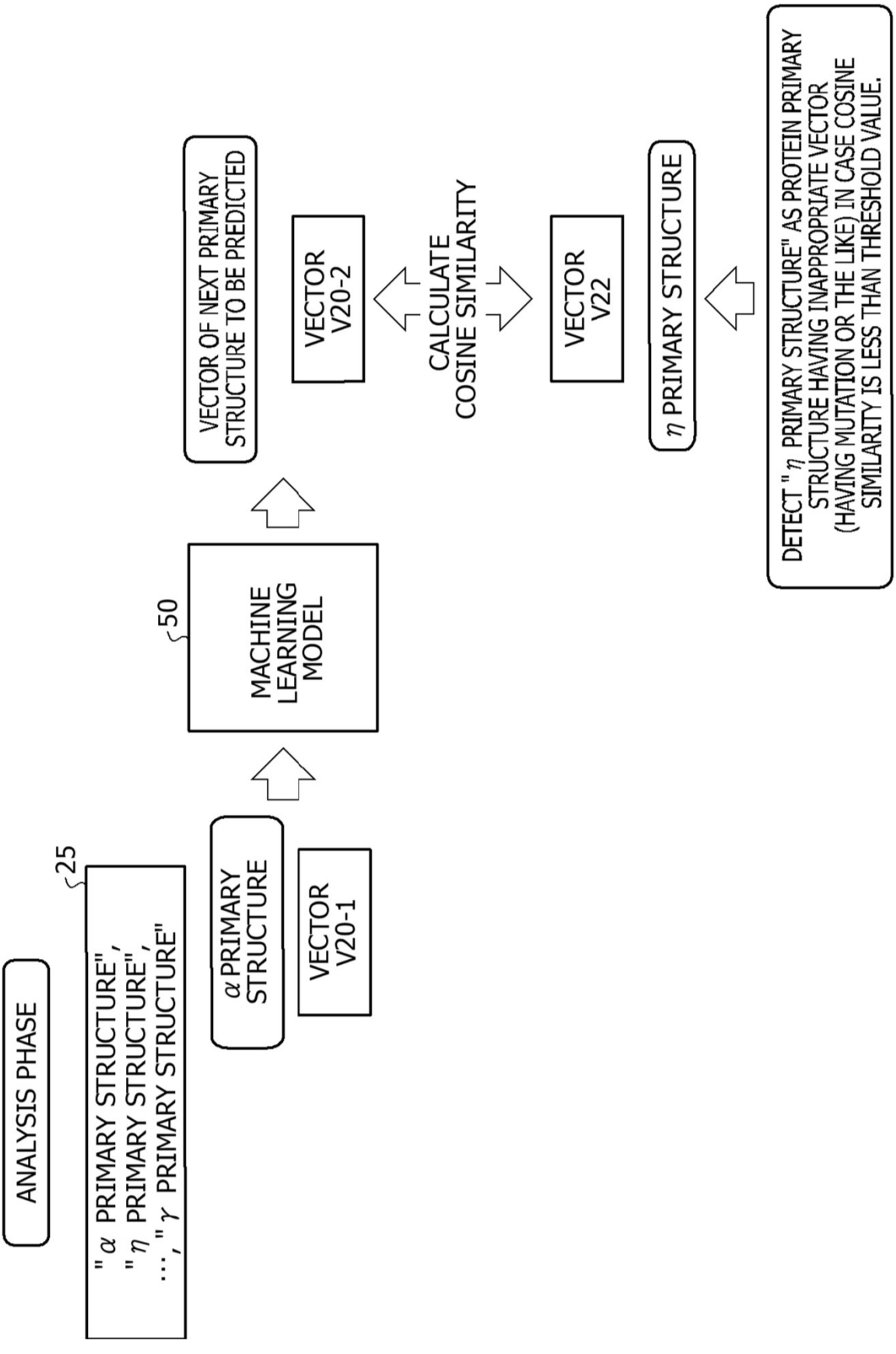
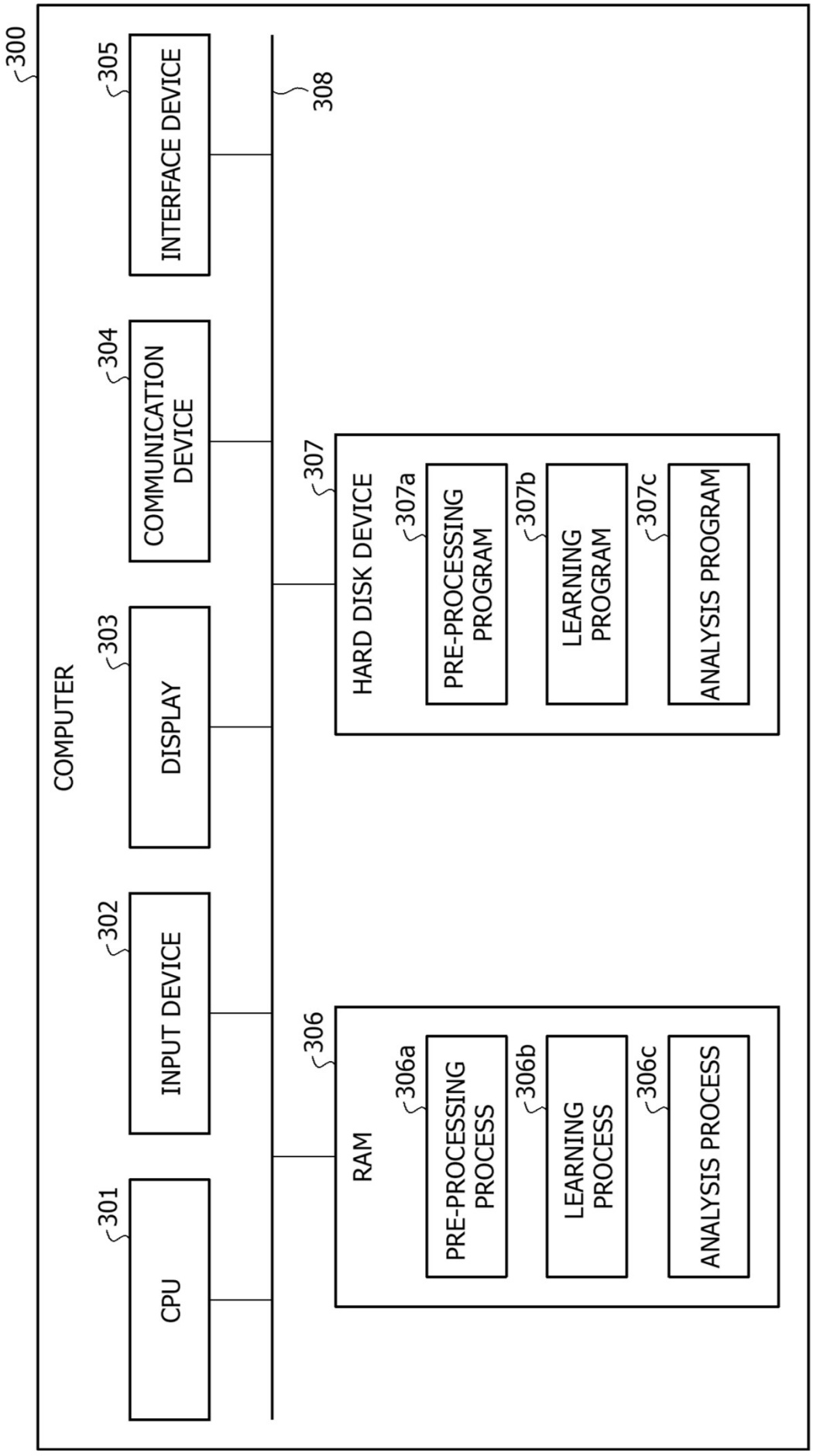

ANALYSIS PHASE
25
"α PRIMARY STRUCTURE",
"η PRIMARY STRUCTURE",
...,"γ PRIMARY STRUCTURE"
α PRIMARY STRUCTURE
VECTOR V20-1
50
MACHINE LEARNING MODEL
VECTOR OF NEXT PRIMARY STRUCTURE TO BE PREDICTED
VECTOR V20-2
CALCULATE COSINE SIMILARITY
VECTOR V22
η PRIMARY STRUCTURE
DETECT "η PRIMARY STRUCTURE" AS PROTEIN PRIMARY STRUCTURE HAVING INAPPROPRIATE VECTOR (HAVING MUTATION OR THE LIKE) IN CASE COSINE SIMILARITY IS LESS THAN THRESHOLD VALUE.

FIG. 11


300
COMPUTER
301
CPU
302
INPUT DEVICE
303
DISPLAY
304
COMMUNICATION DEVICE
305
INTERFACE DEVICE
306
RAM
306a
PRE-PROCESSING PROCESS
306b
LEARNING PROCESS
306c
ANALYSIS PROCESS
307
HARD DISK DEVICE
307a
PRE-PROCESSING PROGRAM
307b
LEARNING PROGRAM
307c
ANALYSIS PROGRAM
308