

(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
2018年3月8日(08.03.2018)



(10) 国際公開番号

WO 2018/043598 A1

(51) 国際特許分類:
G06F 17/27 (2006.01) G06F 17/30 (2006.01)

(21) 国際出願番号: PCT/JP2017/031250

(22) 国際出願日: 2017年8月30日(30.08.2017)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:
特願 2016-173017 2016年9月5日(05.09.2016) JP

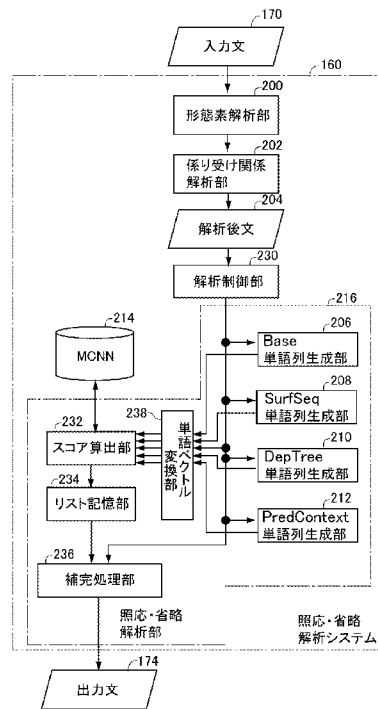
(71) 出願人: 国立研究開発法人情報通信研究機構 (NATIONAL INSTITUTE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY) [JP/JP]; 〒1848795

東京都小金井市貫井北町4-2-1 Tokyo (JP).

(72) 発明者: 飯田 龍(IIDA, Ryu); 〒1848795 東京都小金井市貫井北町4-2-1 国立研究開発法人情報通信研究機構内 Tokyo (JP). 鳥澤 健太郎(TORISAWA, Kentaro); 〒1848795 東京都小金井市貫井北町4-2-1 国立研究開発法人情報通信研究機構内 Tokyo (JP). クルンカライカナサイ(KRUENGKRAI, Canasai); 〒1848795 東京都小金井市貫井北町4-2-1 国立研究開発法人情報通信研究機構内 Tokyo (JP). 呉鍾勲(OH, Jonghoon); 〒1848795 東京都小金井市貫井北町4-2-1 国立研究開発法人情報通信研究機構内 Tokyo (JP). クロエツエージュリアン(KLOETZER, Julien); 〒1848795 東京

(54) Title: CONTEXT ANALYSIS DEVICE AND COMPUTER PROGRAM THEREFOR

(54) 発明の名称: 文脈解析装置及びそのためのコンピュータプログラム



160 Anaphor/omission analysis system
170 Input text
174 Output text
200 Morpheme analysis unit
202 Dependency relationship analysis unit
204 Post-analysis text
206 Base word string generation unit
208 SurfSeq word string generation unit
210 DepTree word string generation unit
212 PredContext word string generation unit
216 Anaphor/omission analysis unit
230 Analysis control unit
232 Score calculation unit
234 List storage unit
236 Supplement processing unit
238 Word vector conversion unit

(57) Abstract: [Problem] To provide a context analysis device that is able to utilize contextual features comprehensively and efficiently to perform context analysis with high precision. [Solution] A context analysis device 160 that includes: an analysis control unit 230 that detects a predicate from which a subject or the like has been omitted, and a supplement candidate therefor; and an anaphor/omission analysis unit 216 that determines a word that should be supplemented. The anaphor/omission analysis unit 216 includes: word vector generation units 206, 208, 210, and 212, which, for supplement



都小金井市貫井北町 4 - 2 - 1 国立研究開発法人情報通信研究機構内 Tokyo (JP).

(74) 代理人: 清水 敏 (SHIMIZU, Satoshi); 〒5300002
大阪府大阪市北区曽根崎新地二丁目 5 番 3
号 堂島 T S S ビル 4 階 Osaka (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, 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, IR, IS, JO, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, 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))

candidates, generate multiple types of word vectors from a text 204; a learned weighting neural network 214 (or LSTM) which, for each supplement candidate, inputs a word vector and calculates a score indicating a probability that the word vector is the omitted word; and a list storage unit 234 and a supplement processing unit 236, which determine the supplement candidate having the best score. The word vectors at least include multiple word vectors each of which is extracted using character strings in the entire text other than the analysis subject and the candidates. Other words, such as deictic words, can be processed in the same manner.

(57) 要約: 【課題】文脈の特徴を包括的かつ効率的に利用して文脈解析を高精度で行える文脈解析装置を提供する。【解決手段】文脈解析装置 160 は、主語等が省略された述語と、その補完候補を検出する解析制御部 230 と、補完すべき語を決定する照応・省略解析部 216 とを含む。照応・省略解析部 216 は、補完候補について、文 204 から複数種類の単語ベクトルを生成する単語ベクトル生成部 206、208、210 及び 212 と、各補完候補について、単語ベクトルを入力として省略された語である確率を示すスコアを出力するよう学習済の畳み込みニューラルネットワーク 214 (又は LSTM) と、スコアが最も良い補完候補を決定するリスト記憶部 234 及び補完処理部 236 とを含む。単語ベクトルは、各々が、少なくとも、解析対象と候補以外の文全体の文字列を用いて抽出される複数個の単語ベクトルを含む。指示語等、他の語についても同様に処理できる。

明 細 書

発明の名称：

文脈解析装置及びそのためのコンピュータプログラム

技術分野

[0001] この発明は、文脈に基づいて、文中のある単語と特定の関係にある別の単語であって、文の単語列からは明確に判定できない単語を特定する、文脈解析装置に関する。より詳しくは、本発明は、文中の指示語が指す語を特定する照応解析、又は文中で主語が省略されている述語の主語を特定する省略解析等を行うための文脈解析装置に関する。

背景技術

[0002] 自然言語の文中には、省略及び指示語が頻出する。例えば図1に示す例文30を考える。例文30は第1文と第2文とからなる。第2文には、「それ」という指示語（代名詞）42が含まれる。指示語42がどの語を指すかは、文の単語列を見ただけでは判断できない。この場合、「それ」という指示語42は第1文の「モン歴の正月の日付」という表現40を指す。このように、文中に存在する指示語の指す語を特定する処理を「照応解析」と呼ぶ。

[0003] これに対し、図2の例文60を考える。この例文60は、第1文と第2文とからなる。第2文において、「自己診断機能を搭載」という述語の主語は省略されている。この主語の省略箇所76には、第1文の「新型交換機」という単語72が省略されている。同様に、「200システムを設置する予定だ。」という述語の主語も省略されている。この主語の省略箇所74には、第1文の「N社」という単語70が省略されている。このように、主語等の省略を検出し、それを補完する処理を「省略解析」と呼ぶ。以後、照応解析と省略解析とをまとめて「照応・省略解析」と呼ぶ。

[0004] 照応解析において指示語がどの語を指しているか、及び、省略解析において省略箇所に補完されるべき単語が何かは、人間には比較的容易に判断できる。この判断には、それら単語が置かれている文脈に関する情報が活用され

ていると思われる。現実には日本語では多数の指示語及び省略が使用されているが、人間が判断する上では大きな支障は生じない。

[0005] 一方、いわゆる人工知能において、人間とのコミュニケーションをとるために、自然言語処理は欠かせない技術である。自然言語処理の重要な問題として、自動翻訳及び質問応答等が存在する。照応・省略解析の技術は、このような自動翻訳及び質問応答において必須の要素技術である。

[0006] しかし、照応・省略解析の現状の性能が実用レベルに至っているとはいえない。その主な理由は、従来型の照応・省略解析技術は、主に指し先の候補と指し元（代名詞及び省略等）から得られる手がかりを利用しているが、その特徴だけでは照応・省略関係を特定することが困難なためである。

[0007] 例えば、後掲の非特許文献１の照応・省略解析アルゴリズムでは、形態素解析・統語解析等の比較的表層的な手がかりに加えて、代名詞・省略を持つ述語と指し先・補完対象となる表現の意味的な整合性を手がかりとして利用している。例として、述語「食べる」の目的語が省略される場合には、「食べ物」に該当する表現を、整備済みの辞書と照合することで「食べる」の目的語を探索する。又は、大規模な文書データから「食べる」の目的語として頻出する表現を探索し、その表現を省略補完する表現として選択したり機械学習で利用する特徴量として使用したりしている。

[0008] それ以外の文脈的特徴としては、照応・省略解析に関して、指し先の候補と指し元(代名詞や省略等)の間の係り受け構造におけるパス中に出現する機能語等を利用する(非特許文献１)こと、及び係り受け構造のパスから解析に有効な部分構造を抽出して利用する(非特許文献２)こと等が試みられている。

[0009] 図３に示す文９０を例としてこれら従来技術について説明する。図３に示す文９０は、述語１００、１０２、及び１０４を含む。これらのうち、述語１０２（「受けた」）の主語が省略１０６となっている。この省略１０６を補完すべき単語候補として、文９０には単語１１０、１１２、１１４及び１１６が存在する。これらのうち、単語１１２（「政府」）が省略１０６を補

完すべき単語である。この単語を自然言語処理においてどのように決定すべきかが問題となる。通常、この単語の推定には機械学習による判別器を用いる。

[0010] 図4を参照して、非特許文献1は、述語とその述語の主語の省略を補完すべき単語候補との間の係り受けパス中の機能語・記号を文脈的な特徴として利用している。そのために従来は、入力文に対する形態素解析及び構文解析を行う。例えば「政府」と省略箇所（「 ϕ 」で示す。）の係り受けパスを考えた場合、非特許文献1では、「が」、「、」、「た」、「を」、「て」、「いる」、「。」、という機能語を素性に利用した機械学習により判別する。

[0011] 一方、非特許文献2では、事前に抽出した文の部分構造から分類に寄与する部分木を獲得し、その係り受けパスを一部抽象化することによって素性の抽出に用いている。例えば図5に示すように、「〈名詞〉が」→「〈動詞〉」という部分木が省略補完に有効という情報が事前に獲得される。

[0012] 文脈的特徴の別の利用方法として、2つの述語で主語が同じか否かを分類する主語共有認識という課題を見出し、それを解くことで得られる情報を使用する手法も存在する（非特許文献3）。この手法によれば、主語を共有する述語集合の中で主語を伝播させることで省略解析の処理を実現する。この手法では述語間の関係が文脈的特徴として利用されている。

[0013] このように、指し先及び指し元の出現文脈を手がかりとして利用しなければ、照応・省略解析の性能向上は難しいと思われる。

先行技術文献

非特許文献

[0014] 非特許文献1: Ryu Iida, Massimo Poesio. A Cross-Lingual ILP Solution to Zero Anaphora Resolution. The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT2011), pp.804-813.2011.

非特許文献2: Ryu Iida, Kentaro Inui, Yuji Matsumoto. Exploiting

Syntactic Patterns as Clues in Zero-Anaphora Resolution. 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING/ACL), pp.625-632. 2006.

非特許文献3: Ryu Iida, Kentaro Torisawa, Chikara Hashimoto, Jong-Hoon Oh, Julien Kloetzer. Intra-sentential Zero Anaphora Resolution using Subject Sharing Recognition. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp.2179-2189, 2015.

非特許文献4: Hiroki Ouchi, Hiroyuki Shindo, Kevin Duh, and Yuj i Matsumoto. 2015. Joint case argument identification for Japanese predicate argument structure analysis. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pages 961-970.

非特許文献5: Ilya Sutskever, Oriol Vinyals, Quoc Le, Sequence to Sequence Learning with Neural Networks, NIPS 2014.

発明の概要

発明が解決しようとする課題

[0015] このように照応・省略解析の性能が向上しない理由として、文脈情報の利用方法に改善の余地があることが挙げられる。既存の解析技術で文脈情報を利用する際には、利用する文脈的特徴を事前に研究者の内省に基づいて取捨選択するという方法が採用されている。しかしこうした方法では、文脈により表される重要な情報が捨てられている可能性が否定できない。そうした問題を解決するためには、重要な情報が捨てられないような方策を採るべきである。しかしそうした問題意識は従来の研究には見ることができず、文脈情報を活かすためにどのような方法を採用すればよいかもよく分かっていなかった。

[0016] それゆえにこの発明の目的は、文脈的特徴を包括的かつ効率的に利用することによって、文中の照応・省略解析等の文解析を高精度で行うことができる文脈解析装置を提供することである。

課題を解決するための手段

[0017] 本発明の第1の局面に係る文脈解析装置は、文の文脈中で、ある単語と一定の関係を持つ別の単語であって、文だけからは当該ある単語と関係を持つことが明確でない別の単語を特定する。この文脈解析装置は、文中で、ある単語を解析対象として検出するための解析対象検出手段と、解析対象検出手段により検出された解析対象について、当該解析対象と一定の関係を持つ別の単語である可能性のある単語候補を文中で探索するための候補探索手段と、解析対象検出手段により検出された解析対象について、候補探索手段により探索された単語候補のうちから1つの単語候補を上記した別の単語として決定するための単語決定手段とを含む。単語決定手段は、単語候補の各々について、文と、解析対象と、当該単語候補とによって定まる、複数種類の単語ベクトル群を生成するための単語ベクトル群生成手段と、単語候補の各々について、単語ベクトル群生成手段により生成された単語ベクトル群を入力として、当該単語候補が解析対象と関係する可能性を示すスコアを出力するよう予め機械学習により学習済のスコア算出手段と、スコア算出手段により出力されたスコアが最も良い単語候補を解析対象と一定の関係を持つ単語として特定する単語特定手段とを含む。複数種類の単語ベクトル群は、各々が、少なくとも、解析対象と単語候補以外の文全体の単語列を用いて生成される1又は複数個の単語ベクトルを含む。

[0018] 好ましくは、スコア算出手段は、複数個のサブネットワークを持つニューラルネットワークであり、複数個の単語ベクトルはそれぞれ、ニューラルネットワークに含まれる複数個のサブネットワークに入力される。

[0019] より好ましくは、単語ベクトル群生成手段は、文全体に含まれる単語列を表す単語ベクトルを出力する第1の生成手段、文のうち、ある単語と単語候補とにより分割された複数個の単語列から、それぞれ単語ベクトルを生成し

出力する第2の生成手段、文を構文解析して得られた係り受け木に基づき、単語候補に係る部分木から得られる単語列、ある単語の係り先の部分木から得られる単語列、単語候補とある単語との間の、係り受け木中の係り受けパスから得られる単語列、及び係り受け木中のそれら以外の部分木からそれぞれ得られる単語列、から得られる単語ベクトルの任意の組合せを生成し出力する第3の生成手段、及び、文中においてある単語の前後の単語列よりそれぞれ得られる単語列を表す2つの単語ベクトルを生成し出力する第4の生成手段、の任意の組み合わせを含む。

[0020] 複数のサブネットワークの各々は、畳み込みニューラルネットワークである。または、複数のサブネットワークの各々は、LSTM (Long Short Term Memory) でもよい。

[0021] さらに好ましくは、ニューラルネットワークは、マルチカラム畳み込みニューラルネットワーク (MCNN) を含み、マルチカラム畳み込みニューラルネットワークの各カラムに含まれる畳み込みニューラルネットワークは、それぞれ別々の単語ベクトルを単語ベクトル群生成手段から受けるように接続される。

[0022] MCNNを構成するサブネットワークのパラメータは互いに同一であってもよい。

[0023] 本発明の第2の局面に係るコンピュータプログラムは、上記したいずれかの文脈解析装置の全ての手段としてコンピュータを機能させる。

図面の簡単な説明

[0024] [図1]照応解析を説明するための模式図である。

[図2]省略解析を説明するための模式図である。

[図3]文脈的特徴の利用例を示すための模式図である。

[図4]非特許文献1に開示された従来の技術を説明するための模式図である。

[図5]非特許文献2に開示された従来の技術を説明するための模式図である。

[図6]本発明の第1の実施の形態に係るマルチカラム畳み込みニューラルネットワーク (MCNN) による照応・省略解析システムの構成を示すブロック

図である。

[図7]図6に示すシステムで利用されるSurfSeqベクトルを説明するための模式図である。

[図8]図6に示すシステムで利用されるDepTreeベクトルを説明するための模式図である。

[図9]図6に示すシステムで利用されるPredContextベクトルを説明するための模式図である。

[図10]図6に示すシステムで利用されるMCNNの概略構成を示すブロック図である。

[図11]図10に示すMCNNの機能を説明するための模式図である。

[図12]図6に示す照応・省略解析部を実現するプログラムの制御構造を示すフローチャートである。

[図13]本発明の第1の実施の形態に係るシステムの効果を説明するグラフである。

[図14]本発明の第2の実施の形態に係るマルチカラム(MC)LSTMによる照応・省略解析システムの構成を示すブロック図である。

[図15]第2の実施の形態における省略の差し先の判定を模式的に説明するための図である。

[図16]図6に示すシステムを実現するためのプログラムを実行するコンピュータの外観を示す図である。

[図17]図16に外観を示すコンピュータのハードウェアブロック図である。

発明を実施するための形態

[0025] 以下の説明及び図面では、同一の部品には同一の参照番号を付してある。したがって、それらについての詳細な説明は繰返さない。

[0026] [第1の実施の形態]

<全体構成>

図6を参照して、最初に、本発明の一実施の形態に係る照応・省略解析システム160の全体構成について説明する。

[0027] この照応・省略解析システム 160 は、入力文 170 を受けて形態素解析を行う形態素解析部 200 と、形態素解析部 200 の出力する形態素列に対して係り受け解析をし、係り受け関係を示す情報が付された解析後文 204 を出力する係り受け関係解析部 202 と、解析後文 204 の中で、文脈解析の対象となる指示語及び主語の省略された述語を検出し、それらの指し先候補及び省略された箇所にも補完すべき単語の候補（補完候補）を探索して、それらの組み合わせの各々に対して指し先及び補完候補を 1 つに決定するための処理を行うために以下の各部の制御を行う解析制御部 230 と、指し先候補及び補完候補を決定するよう予め学習済の MCNN 214 と、解析制御部 230 により制御され、MCNN 214 を参照することによって、解析後文 204 に対する照応・省略解析を行って指示語にはその指示する語を示す情報を付し、省略箇所にはそこに補完すべき単語を特定する情報を付して出力文 174 として出力する照応・省略解析部 216 とを含む。

[0028] 照応・省略解析部 216 は、解析制御部 230 から指示語と指し先の組み合わせ、又は主語が省略された述語とその主語の補完候補の組み合わせをそれぞれ受け、後述する Base ベクトル列、SurfSeq ベクトル列、DepTree ベクトル列、及び PredContext ベクトル列を生成するための単語列を文から抽出する Base 単語列抽出部 206、SurfSeq 単語列抽出部 208、DepTree 単語列抽出部 210、及び PredContext 単語列抽出部 212 と、Base 単語列抽出部 206、SurfSeq 単語列抽出部 208、DepTree 単語列抽出部 210 及び PredContext 単語列抽出部 212 からそれぞれ Base 単語列、SurfSeq 単語列、DepTree 単語列、及び PredContext 単語列を受け、これらの単語列をそれぞれ単語ベクトル（単語埋め込みベクトル；Word Embedding Vector）列に変換する単語ベクトル変換部 238 と、MCNN 214 を用いて、単語ベクトル変換部 238 が出力する単語ベクトル列に基づいて、解析制御部 230 から与えられた組み合わせの指し先候補又は補完候補の各々のスコアを算出して出力するスコア算出部 232 と、スコア算出部 232 が出力するスコアを、各指示語及び省略箇所ごとに、指し先候補又は補完候補のリストとして記憶するリスト記

憶部 234 と、リスト記憶部 234 に記憶されたリストに基づき、解析後文 204 内の指示語及び省略箇所の方々について、最もスコアが高い候補を選択して補完し、補完後の文を出力文 174 として出力するための補完処理部 236 とを含む。

[0029] Base単語列抽出部 206 が抽出するBase単語列、SurfSeq単語列抽出部 208 が抽出するSurfSeq単語列、DepTree単語列抽出部 210 が抽出するDepTree単語列、及びPredContext単語列抽出部 212 が抽出するPredContext単語列はいずれも、文全体から抽出される。

[0030] Base単語列抽出部 206 は、解析後文 204 に含まれる、省略補完の対象となる名詞と省略を持つ可能性のある述語の対から単語列を抽出しBase単語列として出力する。ベクトル変換部 238 が、この単語列から単語ベクトル列であるBaseベクトル列を生成する。本実施の形態では、単語の出現順序を保存し、かつ演算量を少なくするために以下の全ての単語ベクトルとして単語埋め込みベクトルを使用する。

[0031] なお、以下の説明では、理解を容易にするために、主語が省略された述語の主語の候補について、その単語ベクトル列の集合を生成する方法を説明する。

[0032] 図7を参照して、図6に示すSurfSeq単語列抽出部 208 が抽出する単語列は、文90中での単語列の出現順序に基づき、文頭から補完候補 250 までの単語列 260、補完候補 250 と述語 102 の間の単語列 262、及び述語 102 の後、文末までの単語列 264 を含む。したがって、SurfSeqベクトル列は3つの単語埋め込みベクトル列として得られる。

[0033] 図8を参照して、DepTree単語列抽出部 210 が抽出する単語列は、文90の係り受け木に基づき、補完候補 250 に係る部分木 280、述語 102 の係り先の部分木 282、補完候補と述語 102 の間の係り受けパス 284、及びその他 286 からそれぞれ得られる単語列を含む。したがってこの例ではDepTreeベクトル列は4つの単語埋め込みベクトル列として得られる。

[0034] 図9を参照して、PredContext単語列抽出部 212 が抽出する単語列は、文

90において、述語102の前の単語列300と、後の単語列302とを含む。したがってこの場合、PredContextベクトル列は2つの単語埋め込みベクトル列として得られる。

[0035] 図10を参照して、本実施の形態では、MCNN214は、第1～第4の畳み込みニューラルネットワーク群360、362、364、366からなるニューラルネットワーク層340と、ニューラルネットワーク層340内の各ニューラルネットワークの出力を線形に連結する連結層342と、連結層342の出力するベクトルに対してSoftmax関数を適用して、補完候補が真の補完候補か否かを0～1の間のスコアで評価し出力するSoftmax層344とを含む。

[0036] ニューラルネットワーク層340は、前述のとおり第1の畳み込みニューラルネットワーク群360、第2の畳み込みニューラルネットワーク群362、第3の畳み込みニューラルネットワーク群364、及び第4の畳み込みニューラルネットワーク群366を含む。

[0037] 第1の畳み込みニューラルネットワーク群360は、Baseベクトルを受ける第1カラムのサブネットワークを含む。第2の畳み込みニューラルネットワーク群362は、3つのSurfSeqベクトル列をそれぞれ受ける第2、第3及び第4カラムのサブネットワークを含む。第3の畳み込みニューラルネットワーク群364は、4つのDepTreeベクトル列をそれぞれ受ける第5、第6、第7、及び第8カラムのサブネットワークを含む。第4の畳み込みニューラルネットワーク群366は、2つのPredContextベクトル列を受ける第9及び第10カラムのサブネットワークを含む。これらサブネットワークは、いずれも畳み込みニューラルネットワークである。

[0038] ニューラルネットワーク層340の各畳み込みニューラルネットワークの出力は連結層342で単純に線形に連結され、Softmax層344への入力ベクトルとなる。

[0039] MCNN214についてその機能をより詳細に説明する。図11に、代表として1つの畳み込みニューラルネットワーク390を示す。ここでは、説

明を分かりやすくするために、畳み込みニューラルネットワーク 390 が、入力層 400、畳み込み層 402、及びプーリング層 404 のみからなっているものとするが、この 3 つの層を複数個備えているものでもよい。

[0040] 入力層 400 には、単語ベクトル変換部 238 が出力した単語ベクトル列 $X_1, X_2, \dots, X_{|t|}$ がスコア算出部 232 を介して入力される。この単語ベクトル列 $X_1, X_2, \dots, X_{|t|}$ は、行列 $T = [X_1, X_2, \dots, X_{|t|}]^T$ として表される。この行列 T に対して、 M 個の素性マップが適用される。素性マップはベクトルであって、各素性マップの要素であるベクトル O は連続する単語ベクトルからなる N グラムに対して f_j ($1 \leq j \leq M$) で示されるフィルタを適用しながら N グラム 410 を移動させることにより計算される。 N は任意の自然数だが、本実施の形態では $N = 3$ とする。すなわち O は次の式により表される。

[0041] [数1]

$$O = f(W_{f_j} \bullet x_{tj:N-1} + b_{f_j}) \quad (1)$$

ただし $\bullet$ は要素ごとに乗算した後、それらの和をとることを表し、 $f(x) = \max(0, x)$ (正規格線形関数) である。また単語ベクトルの要素数を d とすると重み W_{f_j} は $d \times N$ 次元の実数行列であり、バイアス b_{f_j} は実数である。

なお、素性マップの全体にわたり N を等しくしてもよいし、異なるものがあってもよい。 N としては、2、3、4 及び 5 程度が適当であろう。本実施の形態では、重み行列は全ての畳み込みニューラルネットワークにおいて等しくしてある。これらは互いに異なっても良いが、実際、互いに等しくした方が、各重み行列を独立に学習する場合より精度が高くなる。

[0042] この素性マップの各々について、次のプーリング層 404 は、いわゆるマックスプーリングを行う。すなわち、プーリング層 404 は、例えば素性マップ f_M の要素のうち、最大の要素 420 を選択し、要素 430 として取出す。これを素性マップの各々に対して行うことによって、要素 432、…、430 を取出し、これらを f_1 から f_M の順番に接続して連結層 342 にベクトル 442 として出力する。各畳み込みニューラルネットワークからはこのよ

うにして得られたベクトル440、…、442、…、444が連結層342に出力される。連結層342は、ベクトル440、…、442、…、444を単純に線形に連結してSoftmax層344に与える。なお、プーリング層404としてはマックスプーリングを行うものの方が平均値を採用するものよりも精度が高いと言われている。しかし、もちろん平均値を採用するようにしてもよいし、下位の層の性質をよく表現するものであれば、他の代表値を用いるようにしてもよい。

[0043] 図6に示す照応・省略解析部216について説明する。照応・省略解析部216は、メモリ及びプロセッサを含むコンピュータハードウェア及びその上で実行されるコンピュータソフトウェアにより実現される。図12に、そのようなコンピュータプログラムの制御構造をフローチャート形式で示す。

[0044] 図12を参照して、このプログラムは、解析対象である文から指示語又は主語の省略された述語 $cand_i$ とその補完候補である単語 $pred_i$ とのペア $\langle cand_i; pred_i \rangle$ を全て生成するステップ460と、ステップ460で生成されたあるペアに対して、MCNN214を用いてスコアを計算し、メモリにリストとして記憶させるステップ464を、全てのペアに対して実行するステップ462と、ステップ462で算出されたリストをスコア n の降順でソートするステップ466とを含む。なおここでは、ペア $\langle cand_i; pred_i \rangle$ は、ある述語とその補完候補として可能な単語との全ての可能な組み合わせを示す。すなわち、このペアの集合の中には、各述語も、補完候補もそれぞれ複数回現れ得る。

[0045] このプログラムはさらに、繰返し制御変数 i を0に初期化するステップ468と、変数 i の値がリストの要素数より大きいかを比較し、比較が肯定か否かにしたがって制御を分岐させるステップ470と、ステップ470の比較が否定であることに応答して実行され、ペア $\langle cand_i; pred_i \rangle$ のスコアが所定のしきい値より大きいか否かにしたがって制御を分岐させるステップ474と、ステップ474の判定が肯定であることに応答して実行され、述語 $pred_i$ の補完候補が既に補完済か否かにしたがって制御を分岐させるステップ476と、ステップ476の判定が否定であることに応答して、述語 $pred_i$ の省略

されている主語に $cand_i$ を補完するステップ478とを含む。ステップ474のしきい値としては、例えば0.7～0.9程度の範囲とすることが考えられる。

[0046] このプログラムはさらに、ステップ474の判定が否定であること、ステップ476の判定が否定であること、又はステップ478の処理が終了したことに応答して実行され、 $\langle cand_i; pred_i \rangle$ をリストから削除するステップ480と、ステップ480に続き、変数 i の値に1を加算して制御をステップ470に戻すステップ482と、ステップ470の判定が肯定であることに応答して実行され、補完後の文を出力して処理を終了するステップ472とを含む。

[0047] なお、MCNN214の学習は、通常のニューラルネットワークの学習と同様である。ただし、学習データとしては、上記した10個の単語ベクトルを単語ベクトルとして用いること、及び、処理中の述語と補完候補との組み合わせが正しいか否かを示すデータを学習データに付加することが上記実施の形態のような判別時とは異なる。

[0048] <動作>

図6～図12に示す照応・省略解析システム160は以下のように動作する。入力文170が照応・省略解析システム160に与えられると、形態素解析部200が入力文170の形態素解析を行って形態素列を係り受け関係解析部202に与える。係り受け関係解析部202はこの形態素列に対して係り受け解析を行い、係り受け情報が付された解析後文204を解析制御部230に与える。

[0049] 解析制御部230は、解析後文204内の、主語が省略された全ての述語を検索し、各述語に対する補完候補を解析後文204内で探索して、それらの組み合わせの各々について以下の処理を実行する。すなわち、解析制御部230は、処理対象の述語と補完候補の組み合わせを1つ選択し、Base単語列抽出部206、SurfSeq単語列抽出部208、DepTree単語列抽出部210及びPredContext単語列抽出部212に与える。Base単語列抽出部206、Su

rfSeq単語列抽出部208、DepTree単語列抽出部210及びPredContext単語列抽出部212はそれぞれ、解析後文204からBase単語列、SurfSeq単語列、DepTree単語列及びPredContext単語列を抽出し、単語列群として出力する。これら単語列群は、単語ベクトル変換部238により単語ベクトル列に変換されスコア算出部232に与えられる。

[0050] 解析制御部230は、単語ベクトル変換部238からこの単語ベクトル列が出力されると、スコア算出部232に以下の処理を実行させる。スコア算出部232は、Baseベクトル列をMCNN214の第1の畳み込みニューラルネットワーク群360の1つのサブネットワークの入力に与える。スコア算出部232は3つのSurfSeqベクトル列をMCNN214の第2の畳み込みニューラルネットワーク群362の3つのサブネットワークの入力にそれぞれ与える。スコア算出部232はさらに、4つのDepTreeベクトル列を第3の畳み込みニューラルネットワーク群364の4つのサブネットワークに与え、2つのPredContextベクトル列を第4の畳み込みニューラルネットワーク群366の2つのサブネットワークに与える。MCNN214は、これら入力された単語ベクトルに応答して、与えられた単語ベクトル群に対応する述語と補完候補の組が正しい確率に対応するスコアを算出し、スコア算出部232に与える。スコア算出部232は、この述語と補完候補との組み合わせに対し、スコアを組み合わせでリスト記憶部234に与え、リスト記憶部234はこの組み合わせをリストの1項目として記憶する。

[0051] 解析制御部230が上記した処理を全ての述語と補完候補との組み合わせに対して実行すると、リスト記憶部234には全ての述語と補完候補との組み合わせごとにそれらのスコアがリストされている（図12、ステップ460、462、464）。

[0052] 補完処理部236は、リスト記憶部234に記憶されているリストをスコアの降順でソートする（図12、ステップ466）。補完処理部236はリストの先頭から項目を読み出し、全ての項目について処理が完了した場合（ステップ470でYES）、補完後の文を出力して（ステップ472）処理を

終了する。まだ項目が残っている場合（ステップ４７０でＮＯ）、読出された項目のスコアがしきい値より大きいかな否かを判定する（ステップ４７４）。そのスコアがしきい値以下なら（ステップ４７４でＮＯ）、ステップ４８０でその項目をリストから削除し、次の項目に進む（ステップ４８２からステップ４７０）。そのスコアがしきい値より大きければ（ステップ４７４でＹＥＳ）、ステップ４７６でその項目の述語に対する主語が他の補完候補により既に補完済かな否かを判定する（ステップ４７６）。既に補完済（ステップ４７６でＹＥＳ）ならその項目をリストから削除し（ステップ４８０）、次の項目に進む（ステップ４８２からステップ４７０）。その項目の述語に対する主語が補完済でなければ（ステップ４７６でＮＯ）、ステップ４７８でその述語に対する主語の省略箇所、その項目の補完候補を補完する。さらにステップ４８０での項目をリストから削除し、次の項目に進む（ステップ４８２からステップ４７０）。

[0053] こうして、可能な全ての補完が完了すると、ステップ４７０の判定がＹＥＳとなり、ステップ４７２で補完後の文が出力される。

[0054] 以上のように、本実施の形態によれば、従来と異なり、文を構成する全ての単語列を用いて、かつ複数の異なる観点から生成されたベクトルを用いて、述語と補完候補（又は指示語とその指し先候補）の組み合わせが正しいものかな否かを判定する。従来のように人手で単語ベクトルを調整することなく、様々な観点から判定することが可能になり、照応・省略解析の精度を上げることが期待できる。

[0055] 実際、実験により上記実施の形態の考え方による照応・省略解析の精度が従来のもものよりも高くなることが確認できた。その結果を図１３にグラフ形式で示す。この実験では、非特許文献３で使用されたものと同じコーパスを用いた。このコーパスは予め述語とその省略箇所の補完語との対応付けが人手でなされたものである。このコーパスを５個のサブコーパスに分割し、３個を学習データ、１個を開発セット、１個をテストデータとして用いた。このデータを用い、上記した実施の形態にしたがった照応・補完手法と、他の

3種類の比較手法とによって省略箇所の補完処理を行い、その結果を比較した。

[0056] 図13を参照して、グラフ500は上記実施の形態にしたがって行った実験結果のPR曲線である。この実験では、上記した4種類の単語ベクトルを全て用いた。グラフ506は、マルチカラムではなく単カラムの畳み込みニューラルネットワークを用い、文に含まれる全ての単語から単語ベクトルを生成して得た例のPR曲線である。黒四角502及びグラフ504で示されるのは、比較のために、非特許文献4に示されたグローバル最適化方法の結果及び実験により得たPR曲線である。この方法では開発セットが不要であるため、開発セットを含めた4つのサブコーパスを学習に用いた。この方法では主語、目的語、間接目的語について述語－文法項間の関係が得られるが、本実験では文中での主語省略の補完についてだけに関する出力を用いた。非特許文献4に示されたものと同様、10回の独立した試行の結果を平均したものを用いている。さらに、非特許文献3の手法を用いた結果508もグラフ中にxで示す。

[0057] 図13から明らかなように、上記実施の形態による手法によれば他のどの手法のものよりもよいPR曲線が得られ、広い範囲で適合率が高い。したがって、上記したような単語ベクトルの選択方法が従来方法で用いられたものよりも適切に文脈情報を表現していると考えられる。さらに、上記実施の形態による方法によれば、単カラムのニューラルネットワークを用いたものよりも高い適合率が得られた。これはMCNNを用いることにより、再現率を高めることができたことを示す。

[0058] [第2の実施の形態]

<構成>

第1の実施の形態に係る照応・省略解析システム160では、スコア算出部232におけるスコア算出にMCNN214を用いている。しかし、本発明はそのような実施の形態には限定されない。MCNNに代えて、LSTMと呼ばれるネットワークアーキテクチャを構成素とするニューラルネットワ

ークを用いてもよい。以下、LSTMを用いた実施の形態について説明する。

[0059] LSTMは、リカレント型ニューラルネットワークの一種であり、入力系列を記憶しておく能力を持つ。実装上、いろいろな変種があるが、入力の系列と、それに対する出力の系列とを一組とする多数組の学習データで学習し、入力の系列を受けると、それに対する出力の系列を受ける仕組みを実現することができる。この仕組みを用いて英語からフランス語に自動翻訳するシステムがすでに利用されている（非特許文献5）。

[0060] 図14を参照して、この実施の形態でMCNN214に代えて用いられるMCLSTM（マルチカラムLSTM）530は、LSTM層540と、第1の実施の形態の連結層342と同様、LSTM層540内の各LSTMの出力を線形に連結する連結層542と、連結層542の出力するベクトルに対してSoftmax関数を適用して、補完候補が真の補完候補か否かを0～1の間のスコアで評価し出力するSoftmax層544とを含む。

[0061] LSTM層540は、第1のLSTM群550、第2のLSTM群552、第3のLSTM群554、及び第4のLSTM群556を含む。これらはいずれもLSTMからなるサブネットワークを含む。

[0062] 第1のLSTM群550は、第1の実施の形態の第1の畳み込みニューラルネットワーク群360と同様、Baseベクトル列を受ける第1カラムのLSTMを含む。第2のLSTM群552は、第1の実施の形態の第2の畳み込みニューラルネットワーク群362と同様、3つのSurfSeqベクトル列をそれぞれ受ける第2、第3及び第4カラムのLSTMを含む。第3のLSTM群554は、第1の実施の形態の第3の畳み込みニューラルネットワーク群364と同様、4つのDepTreeベクトル列をそれぞれ受ける第5、第6、第7、及び第8カラムのLSTMを含む。第4のLSTM群556は、第1の実施の形態の第4の畳み込みニューラルネットワーク群366と同様、2つのPre dContextベクトル列を受ける第9及び第10のLSTMを含む。

[0063] LSTM層540の各LSTMの出力は連結層542で単純に線形に連結

され、Softmax層544への入力ベクトルとなる。

[0064] ただし、本実施の形態では、各単語ベクトル列は、例えば、出現順序にしたがって単語ごとに生成した単語ベクトルからなるベクトル系列の形で生成される。これらベクトル系列を形成する単語ベクトルは、それぞれ単語の出現順序に従って対応のLSTMに順次与えられる。

[0065] LSTM層540を構成するLSTM群の学習も、第1の実施の形態と同様、MCLSTM530の全体についての、学習データを用いた誤差逆伝播法により行われる。この学習は、ベクトル系列が与えられると、MCLSTM530が、補完候補である単語が真に指し先である確率を出力するように行われる。

[0066] <動作>

この第2の実施の形態に係る照応・省略解析システムの動作は、基本的に第1の実施の形態の照応・省略解析システム160と同様である。LSTM層540を構成する各LSTMへのベクトル列の入力も第1の実施の形態と同様である。

[0067] 手順は第1の実施の形態と同様で、その概略は図12に示されている。違いは、図12のステップ464で、第1の実施の形態のMCNN214（図10）に代えて、図14に示すMCLSTM530を使用する点、及び単語ベクトル列として単語ベクトルからなるベクトル系列を用い、各単語ベクトルを順にMCLSTM530に入力する点である。

[0068] 本実施の形態では、LSTM層540を構成する各LSTMにベクトル系列の各単語ベクトルが入力されるたびに、各LSTMはその内部状態を変え、出力も変わる。ベクトル系列の入力が終了した時点での各LSTMの出力は、それまでに入力されたベクトル系列に応じて定まる。連結層542は、それらの出力を連結してSoftmax層544への入力とする。Softmax層544は、この入力に対するsoftmax関数の結果を出力する。この値は上記したように、ベクトル系列を生成する際の、指示語、又は主語が省略された述語に対する指し先の補完候補が真の指し先候補か否かを示す確率である。ある補完

候補に対して算出されるこの確率が、他の補完候補に対して算出された確率より大きく、かつあるしきい値 θ より大きい場合に、その補完候補が真の指し先候補であると推定する。

[0069] 図15(A)を参照して、例文570において、述語である「受けた」という文言580に対する主語が不明であり、その補完候補として「報告書」「政府」及び「条約」という単語582、584及び586が検出されたものとする。

[0070] 図15(B)に示す様に、単語582、584及び586に対して、それぞれ単語ベクトルを表すベクトル系列600、602、及び604が得られ、これらをMCLSTM530への入力として与える。その結果、MCLSTM530の出力として、ベクトル系列600、602、及び604に対してそれぞれ0.5、0.8、及び0.4という値が得られたものとする。これらの最大値は0.8である。また、この0.8という値がしきい値 θ 以上であれば、ベクトル系列602に対応する単語584、すなわち「政府」が「受けた」の主語であると推定される。

[0071] 図12に示すように、こうした処理を対象となる文の中の全ての指示語、又は主語が省略された述語と、それらの指し先候補とのペアに対して実行していくことで、対象文の解析が行われる。

[0072] [コンピュータによる実現]

上記第1及び第2の実施の形態に係る照応・省略解析システムは、コンピュータハードウェアと、そのコンピュータハードウェア上で実行されるコンピュータプログラムとにより実現できる。図16はこのコンピュータシステム630の外観を示し、図17はコンピュータシステム630の内部構成を示す。

[0073] 図16を参照して、このコンピュータシステム630は、メモリポート652及びDVD(Digital Versatile Disc)ドライブ650を有するコンピュータ640と、いずれもコンピュータ640に接続されたキーボード646、マウス648、及びモニタ642とを含む。

[0074] 図17を参照して、コンピュータ640は、メモリポート652及びDVDドライブ650に加えて、CPU（中央処理装置）656と、CPU656、メモリポート652及びDVDドライブ650に接続されたバス666と、ブートプログラム等を記憶する読出専用メモリ（ROM）658と、バス666に接続され、プログラム命令、システムプログラム及び作業データ等を記憶するランダムアクセスメモリ（RAM）660と、ハードディスク654とを含む。コンピュータシステム630はさらに、他端末との通信を可能とするネットワーク668への接続を提供するネットワークインターフェイス（I/F）644を含む。

[0075] コンピュータシステム630を上記した実施の形態に係る照応・省略解析システムの各機能部として機能させるためのコンピュータプログラムは、DVDドライブ650又はメモリポート652に装着されるDVD662又はリムーバブルメモリ664に記憶され、さらにハードディスク654に転送される。又は、プログラムはネットワーク668を通じてコンピュータ640に送信されハードディスク654に記憶されてもよい。プログラムは実行の際にRAM660にロードされる。DVD662から、リムーバブルメモリ664から又はネットワーク668を介して、直接にRAM660にプログラムをロードしてもよい。

[0076] このプログラムは、コンピュータ640を、上記実施の形態に係る照応・省略解析システムの各機能部として機能させるための複数の命令からなる命令列を含む。コンピュータ640にこの動作を行わせるのに必要な基本的機能のいくつかはコンピュータ640上で動作するオペレーティングシステム若しくはサードパーティのプログラム又はコンピュータ640にインストールされる、ダイナミックリンク可能な各種プログラミングツールキット又はプログラムライブラリにより提供される。したがって、このプログラム自体はこの実施の形態のシステム及び方法を実現するのに必要な機能全てを必ずしも含まなくてよい。このプログラムは、命令の内、所望の結果が得られるように制御されたやり方で適切な機能又はプログラミングツールキット又は

プログラムライブラリ内の適切なプログラムを実行時に動的に呼出すことにより、上記したシステムとしての機能を実現する命令のみを含んでいればよい。もちろん、プログラムのみで必要な機能を全て提供してもよい。

[0077] [可能な変形例]

上記実施の形態では、日本語に対する照応・解析処理を扱っている。しかし本発明はそのような実施の形態には限定されない。文全体の単語列を使い、複数の観点で単語ベクトル群を作成するという考え方は、どのような言語にも適用できる。したがって、指示語及び省略が頻発する他の言語（中国語、韓国語、イタリア語、スペイン語）等についても本発明を適用できると考えられる。

[0078] また、上記実施の形態では、文全体の単語列を用いた単語ベクトル列として4種類を用いているが、単語ベクトル列としてはこの4種類に限定されるわけではない。異なる観点から文全体の単語列を用いて作成する単語ベクトル列であればどのようなものでも利用できる。さらに、文全体の単語列を用いるものを少なくとも2種類用いるのであれば、それら以外に文の一部の単語列を用いる単語ベクトル列を加えて使用してもよい。また、単なる単語列だけではなく、それらの品詞情報まで含めた単語ベクトル列を用いるようにしても良い。

今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味及び範囲内での全ての変更を含む。

産業上の利用可能性

[0079] 本発明は、人間とのインタラクションが必要な装置及びサービス全般に適用可能であり、さらに、人間の発話を解析することにより様々な装置及びサービスにおける人間とのインターフェイスを改善するための装置及びサービスに利用できる。

符号の説明

[0080]	9 0	文	
	1 0 0、1 0 2、1 0 4	述語	
	1 0 6	省略	
	1 1 0、1 1 2、1 1 4、1 1 4	単語	
	1 6 0	照応・省略解析システム	
	1 7 0	入力文	
	1 7 4	出力文	
	2 0 0	形態素解析部	
	2 0 2	係り受け関係解析部	
	2 0 4	解析後文	
	2 0 6	Base単語列抽出部	
	2 0 8	SurfSeq単語列抽出部	
	2 1 0	DepTree単語列抽出部	
	2 1 2	PredContext単語列抽出部	
	2 1 4	M C N N	
	2 1 6	照応・省略解析部	
	2 3 0	解析制御部	
	2 3 2	スコア算出部	
	2 3 4	リスト記憶部	
	2 3 6	補完処理部	
	2 3 8	単語ベクトル変換部	
	2 5 0	補完候補	
	2 6 0、2 6 2、2 6 4、3 0 0、3 0 2	単語列	
	2 8 0、2 8 2	部分木	
	2 8 4	係り受けパス	
	3 4 0	ニューラルネットワーク層	
	3 4 2、5 4 2	連結層	
	3 4 4、5 4 4	Softmax層	

3 6 0	第 1 の畳み込みニューラルネットワーク群
3 6 2	第 2 の畳み込みニューラルネットワーク群
3 6 4	第 3 の畳み込みニューラルネットワーク群
3 6 6	第 4 の畳み込みニューラルネットワーク群
3 9 0	畳み込みニューラルネットワーク
4 0 0	入力層
4 0 2	畳み込み層
4 0 4	プーリング層
5 3 0	M C L S T M
5 4 0	L S T M 層
5 5 0	第 1 の L S T M 群
5 5 2	第 2 の L S T M 群
5 5 4	第 3 の L S T M 群
5 5 6	第 4 の L S T M 群
6 0 0、6 0 2、6 0 4	ベクトル系列

請求の範囲

[請求項1]

テキスト文の文脈中で、ある単語と一定の関係を持つ別の単語であって、前記テキスト文だけでは当該ある単語と前記関係を持つことが明確でない前記別の単語を特定する文脈解析装置であって、

前記テキスト文中で、前記ある単語を解析対象として検出するための解析対象検出手段と、

前記解析対象検出手段により検出された解析対象について、当該解析対象と前記一定の関係を持つ前記別の単語である可能性のある単語候補を前記テキスト文中で探索するための候補探索手段と、

前記解析対象検出手段により検出された解析対象について、前記候補探索手段により探索された単語候補のうちから一つの単語候補を前記別の単語として決定するための単語決定手段とを含み、

前記単語決定手段は、

前記単語候補の各々について、前記テキスト文と、前記解析対象と、当該単語候補とによって定まる、複数種類の単語ベクトル群を生成するための単語ベクトル群生成手段と、

前記単語候補の各々について、前記単語ベクトル群生成手段により生成された単語ベクトル群を入力として、当該単語候補が前記解析対象と関係する可能性を示すスコアを出力するよう予め機械学習により学習済のスコア算出手段と、

前記スコア算出手段により出力されたスコアが最も良い単語候補を前記解析対象と前記一定の関係を持つ単語として特定する単語特定手段とを含み、

前記複数種類の単語ベクトル群は、各々が、少なくとも、前記解析対象と前記単語候補以外の前記テキスト文全体の単語列を用いて生成される1又は複数個の単語ベクトルを含む、文脈解析装置。

[請求項2]

前記スコア算出手段は、複数個のサブネットワークを持つニューラルネットワークであり、

前記 1 又は複数個の単語ベクトルはそれぞれ、前記ニューラルネットワークに含まれる前記複数個のサブネットワークに入力される、請求項 1 に記載の文脈解析装置。

[請求項3] 前記複数個のサブネットワークの各々は、畳み込みニューラルネットワークである、請求項 2 に記載の文脈解析装置。

[請求項4] 前記複数個のサブネットワークの各々は、L S T Mである、請求項 2 に記載の文脈解析装置。

[請求項5] 前記単語ベクトル群生成手段は、
前記テキスト文全体に含まれる単語列を表す単語ベクトル列を出力する第 1 の生成手段、

前記テキスト文のうち、前記ある単語と前記単語候補とにより分割された複数個の単語列から、それぞれ単語ベクトル列を生成し出力する第 2 の生成手段、

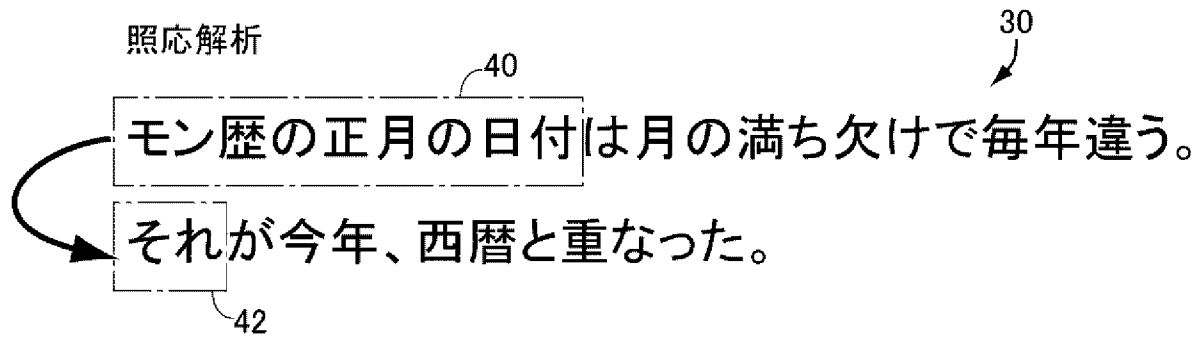
前記テキスト文を構文解析して得られた係り受け木に基づき、前記単語候補に係る部分木から得られる単語列、前記ある単語の係り先の部分木から得られる単語列、前記単語候補と前記ある単語との間の、前記係り受け木中の係り受けパスから得られる単語列、及び前記係り受け木中のそれら以外の部分木からそれぞれ得られる単語列、から得られる単語ベクトル列の任意の組合せを生成し出力する第 3 の生成手段、及び、

前記テキスト文中において前記ある単語の前後の単語列よりそれぞれ得られる単語列を表す 2 つの単語ベクトル列を生成し出力する第 4 の生成手段、

の任意の組み合わせを含む、請求項 1 ～請求項 4 のいずれかに記載の文脈解析装置。

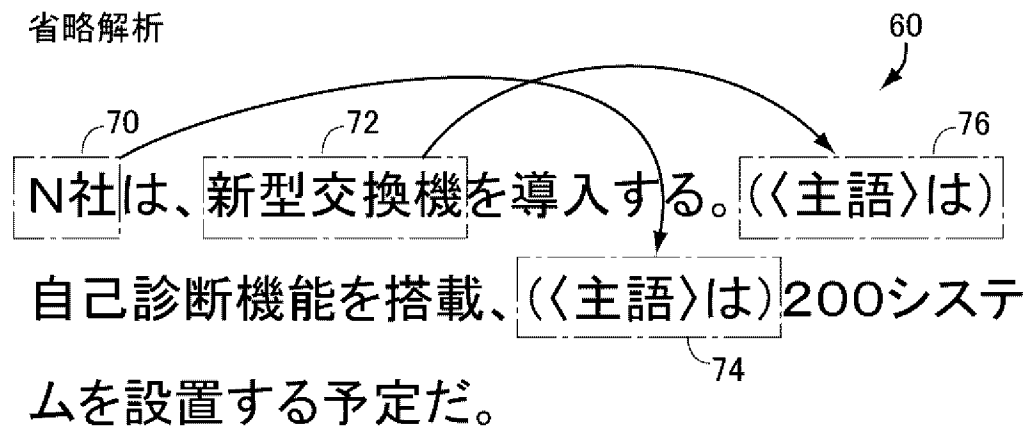
[請求項6] 請求項 1 から請求項 5 の何れかに記載の文脈解析装置としてコンピュータを機能させる、コンピュータプログラム。

[図1]

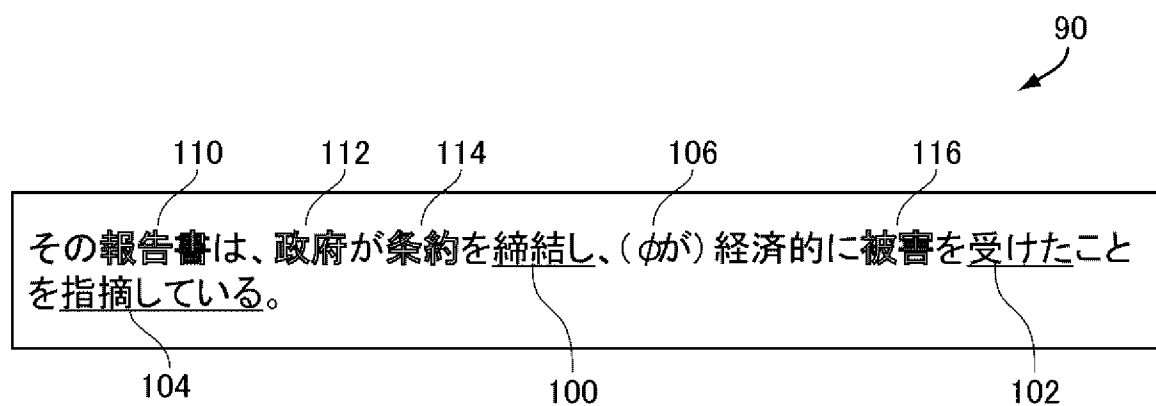


[図2]

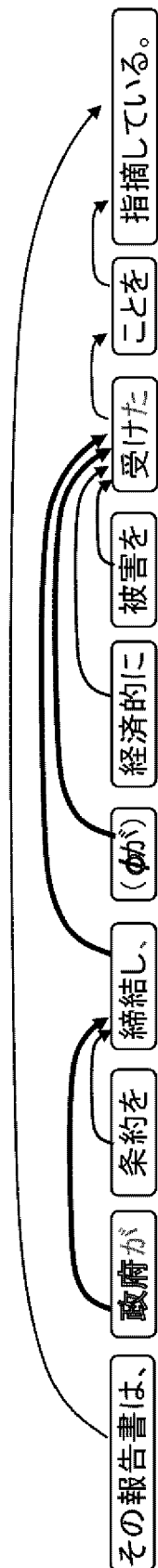
省略解析



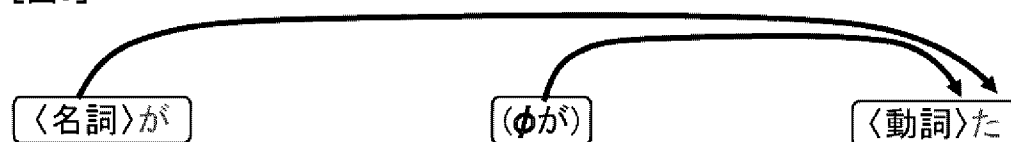
[図3]



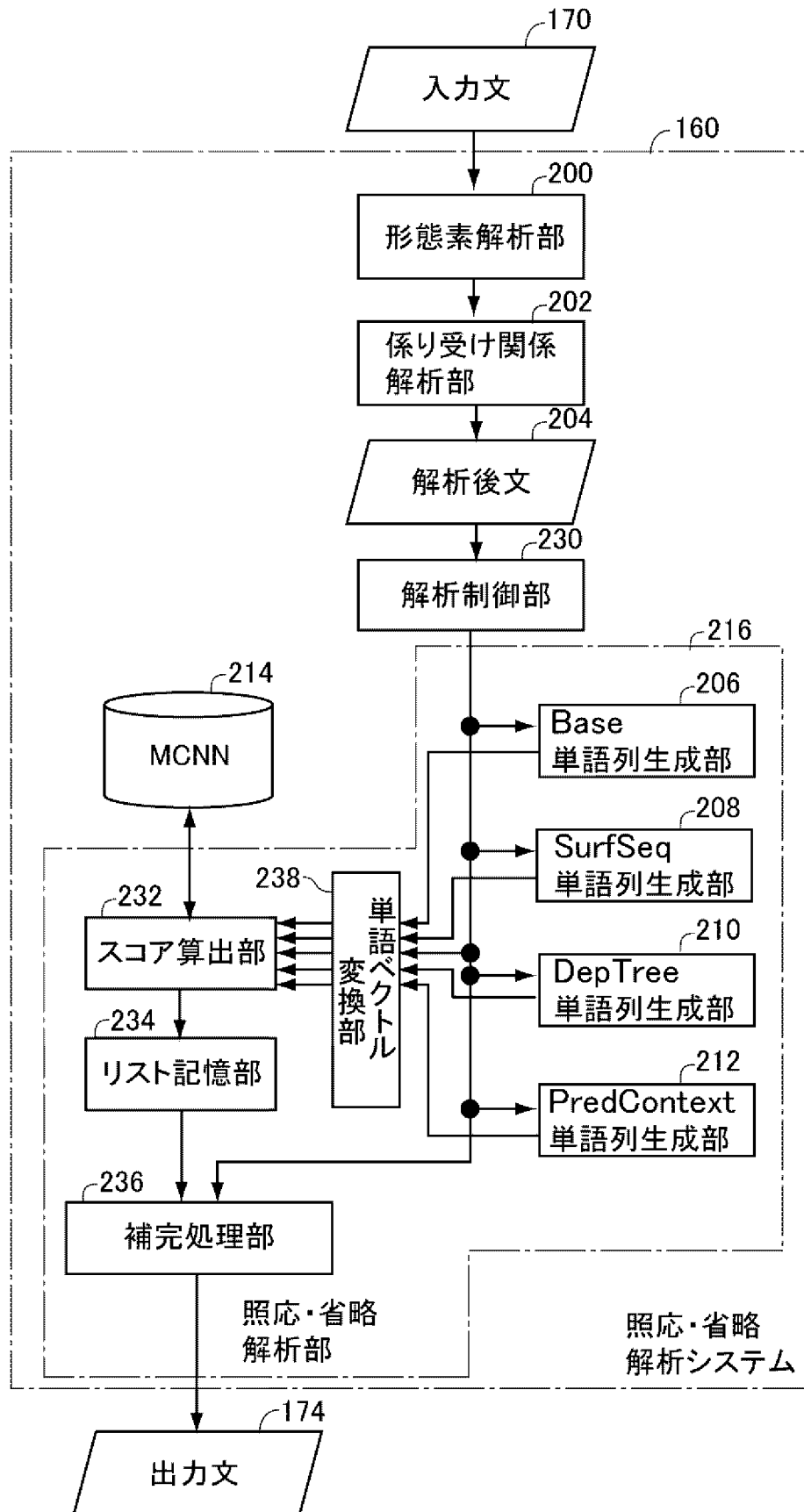
[図4]



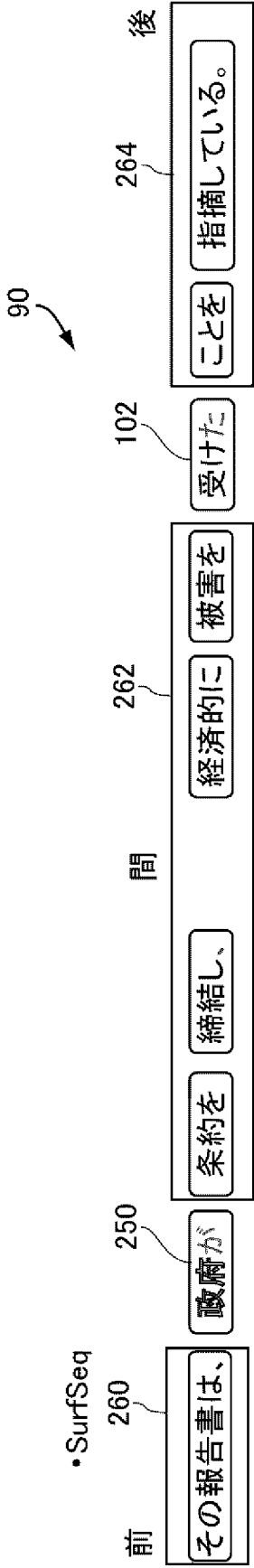
[図5]



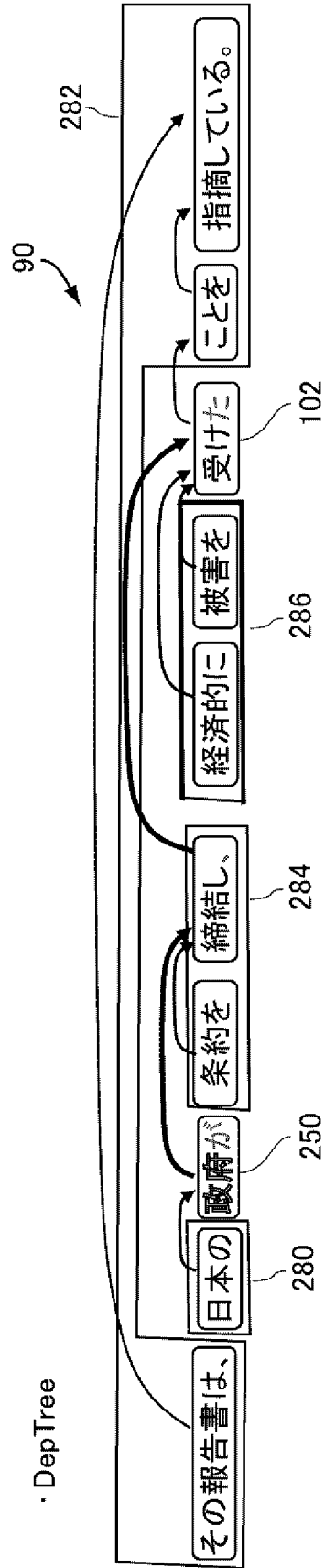
[図6]



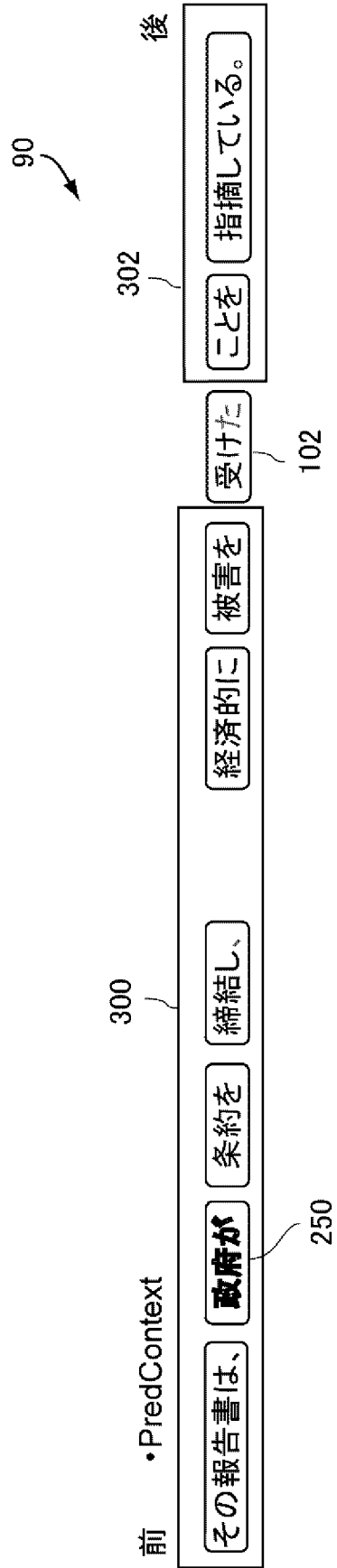
[図7]



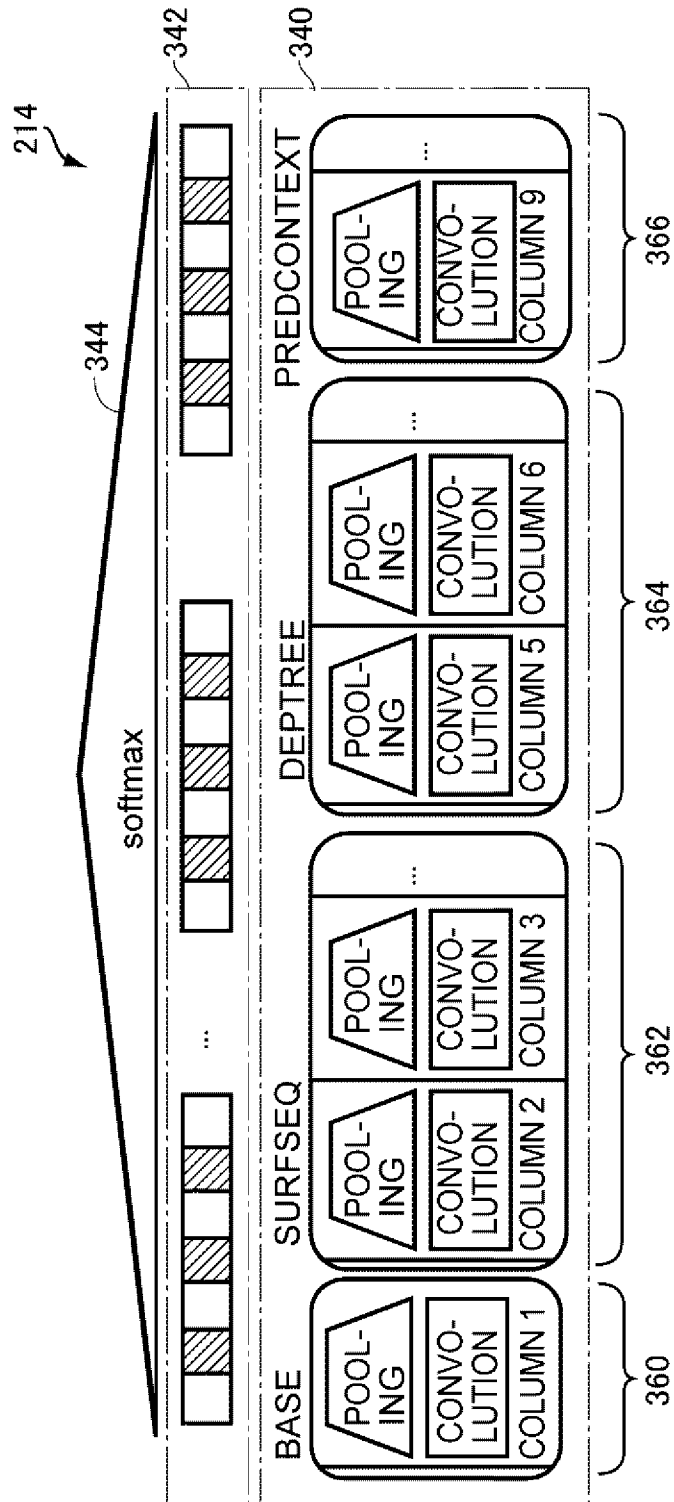
[図8]



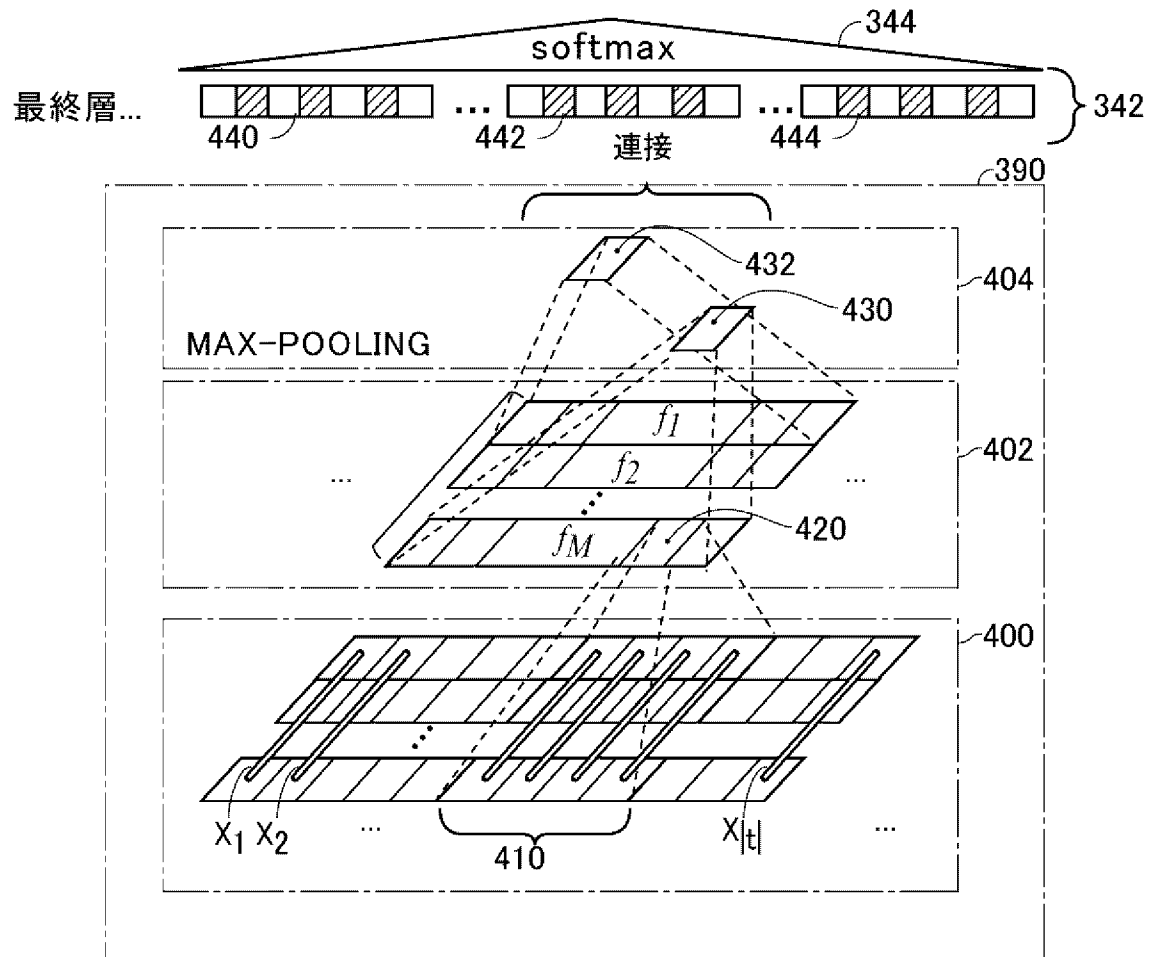
[図9]



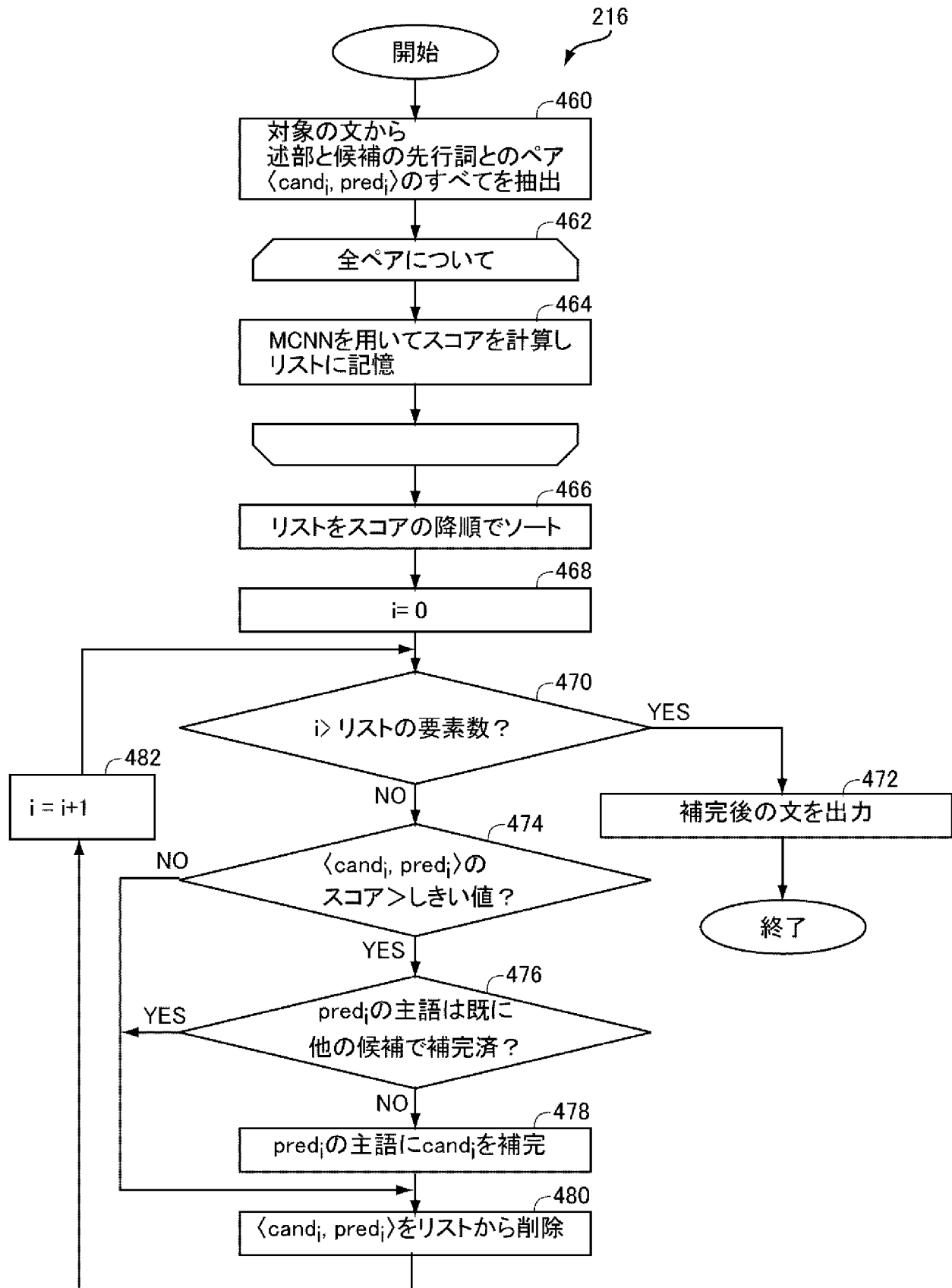
[図10]



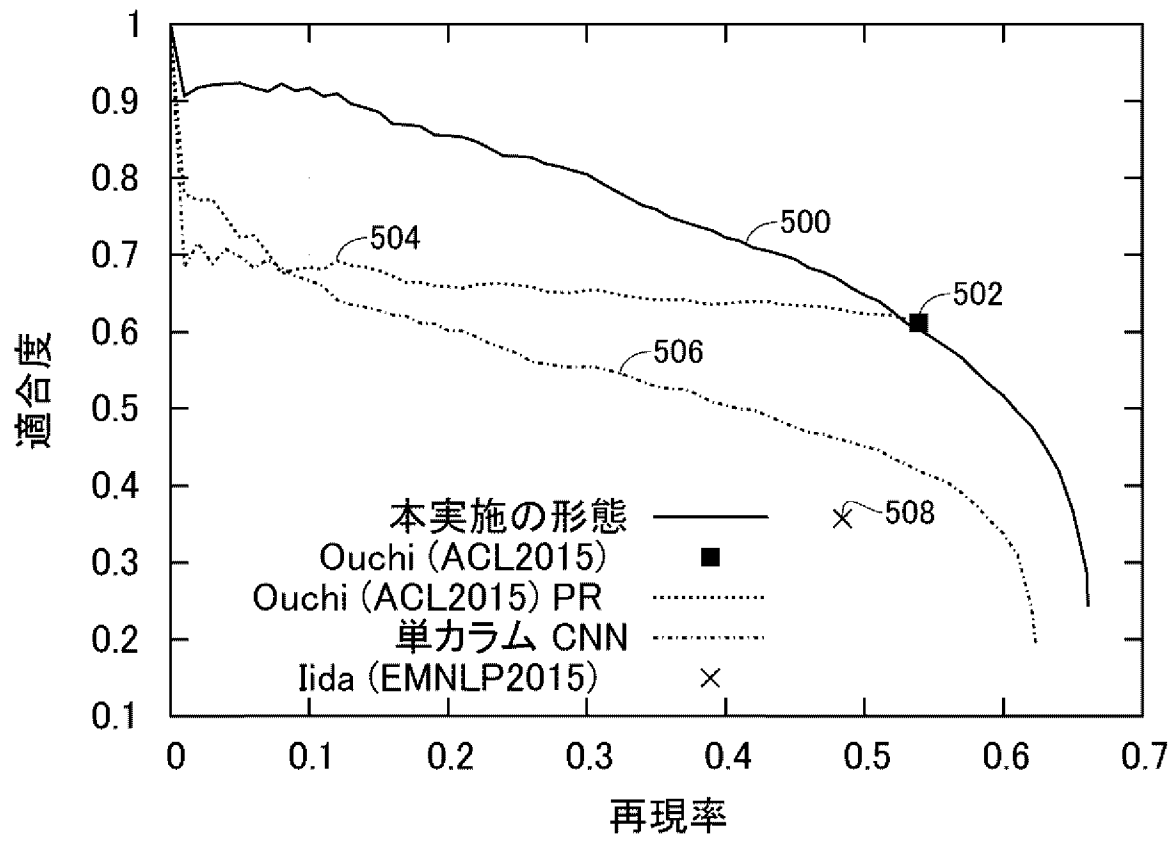
[図11]



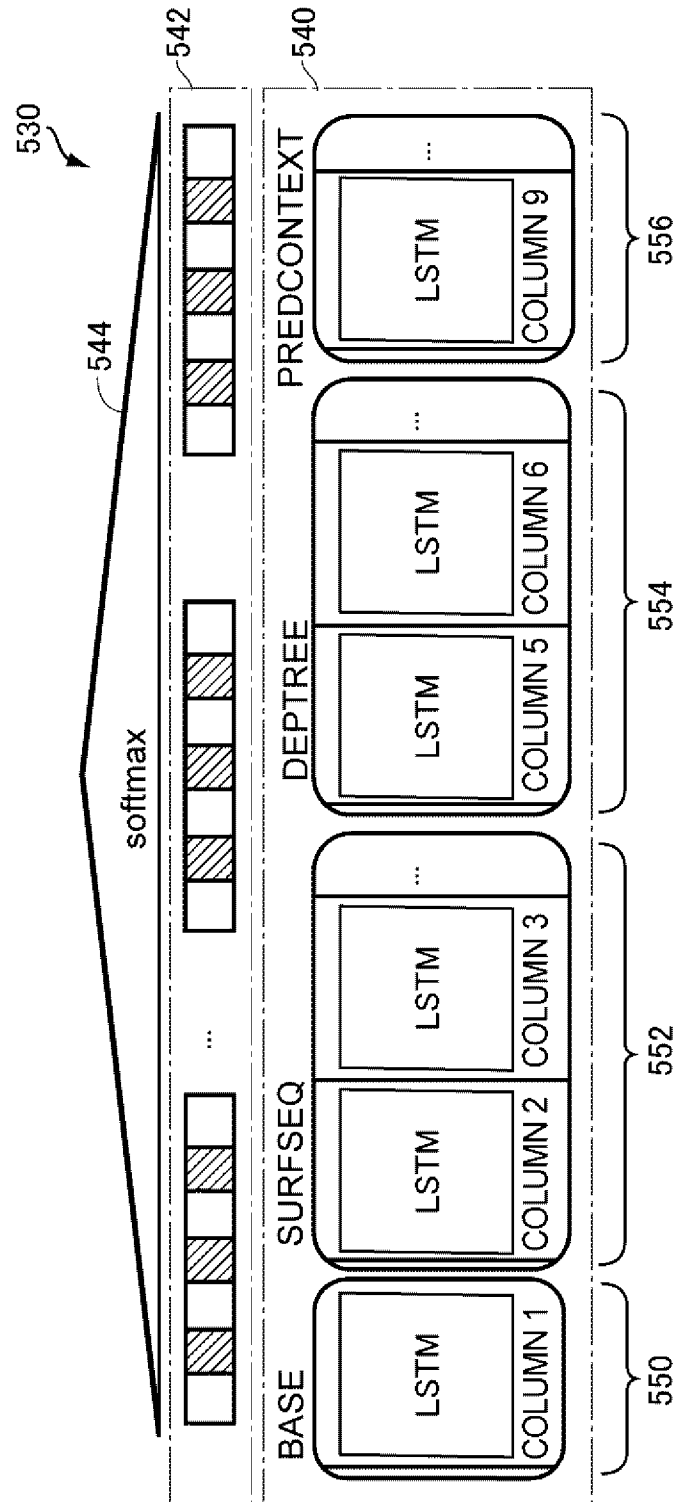
[図12]



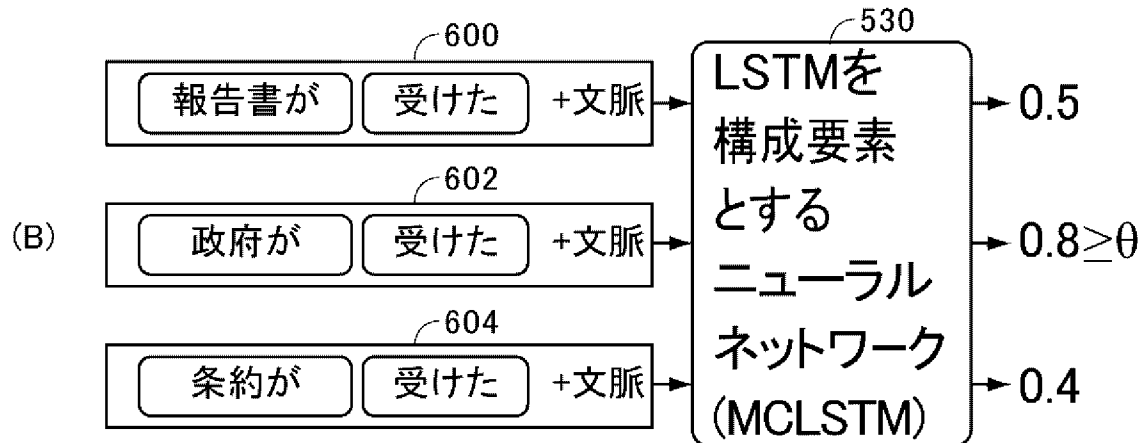
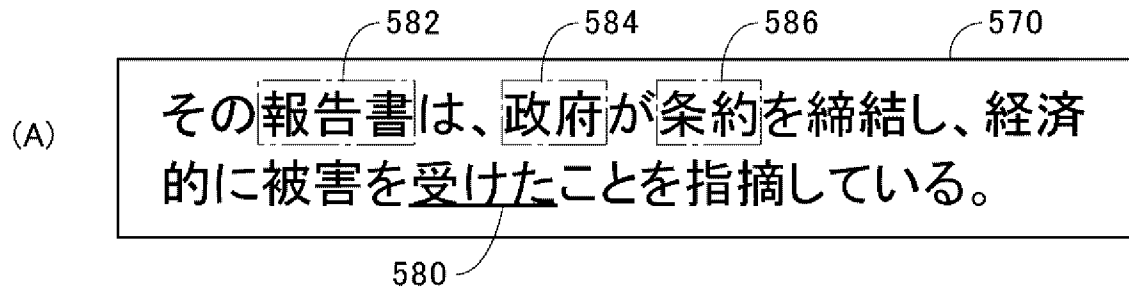
[図13]



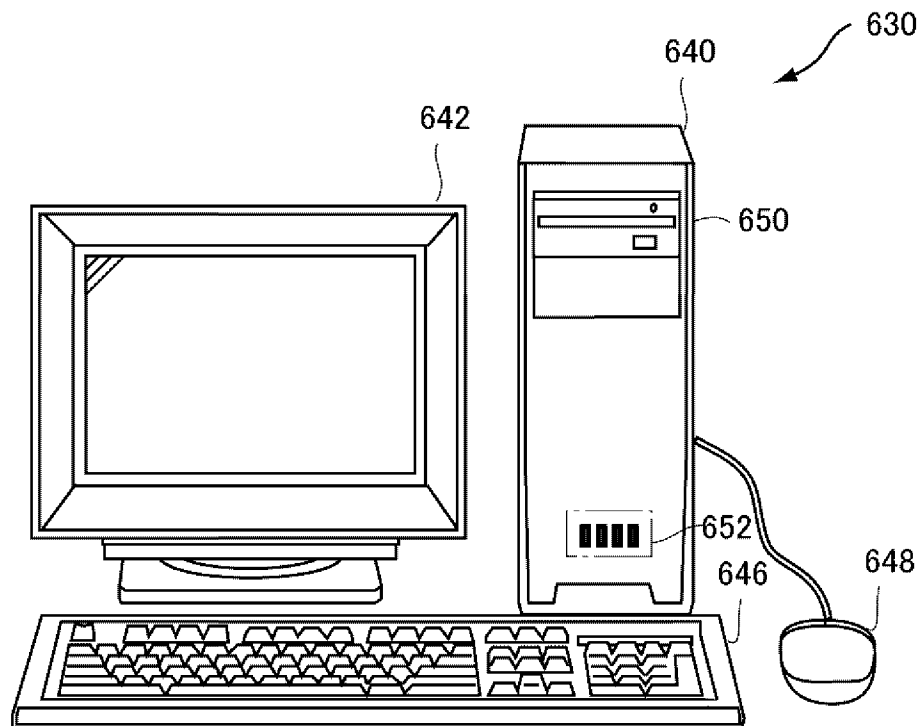
[図14]



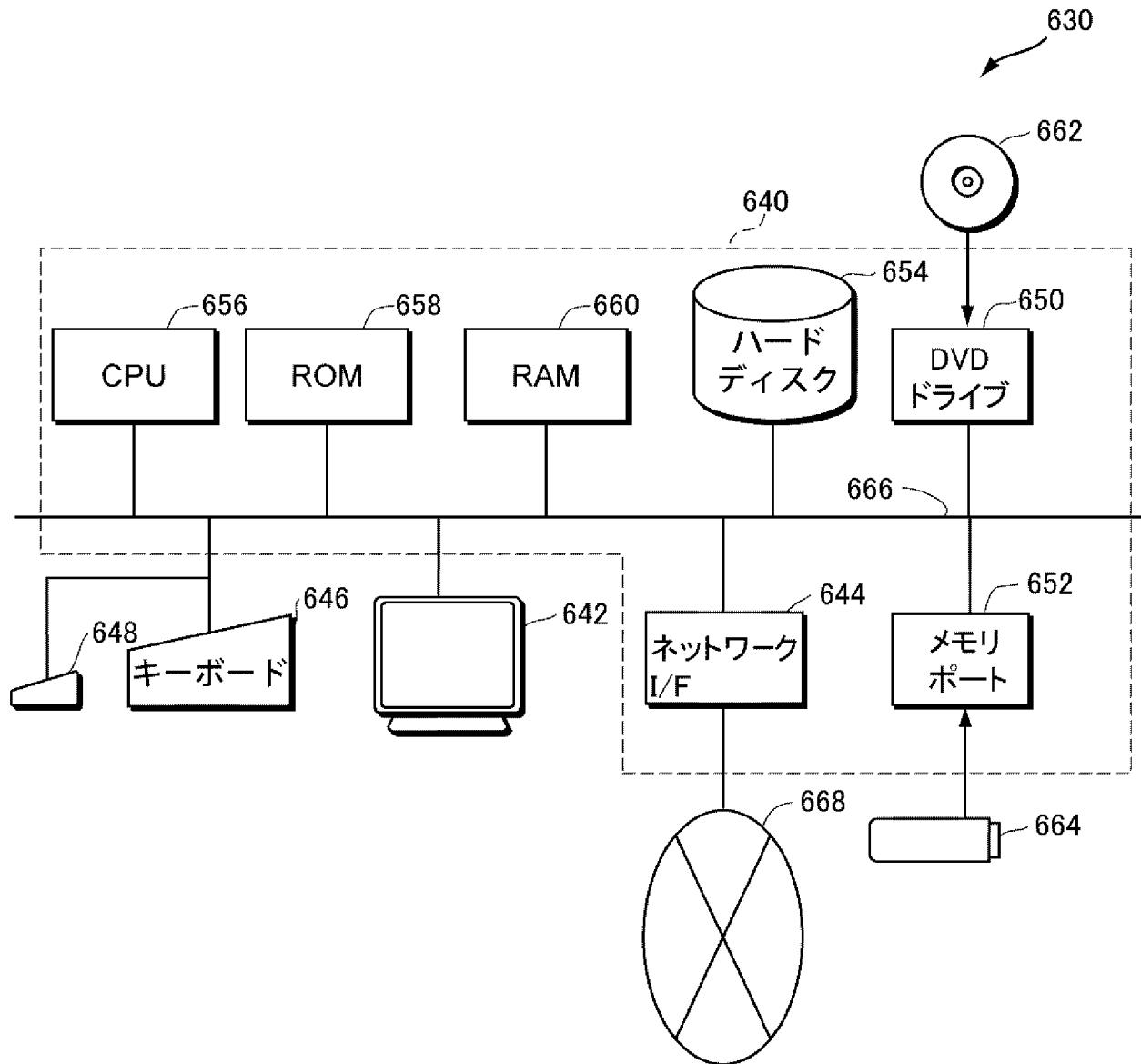
[図15]



[図16]



[図17]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2017/031250

A. CLASSIFICATION OF SUBJECT MATTER

G06F17/27(2006.01)i, G06F17/30(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F17/27, G06F17/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Japanese Published Examined Utility Model Applications 1922-1996

Japanese Published Unexamined Utility Model Applications 1971-2017

Japanese Examined Utility Model Registrations 1996-2017

Japanese Registered Utility Model Specifications 1994-2017

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	江原暉将 外 1 名, 確率モデルによる省略主語の補完, NHK 技研 R&D, 15 February 1997, no. 44, pp. 12-28 (EHARA, Terumasa et al., Zero-subject resolution by probabilistic model, NHK STRL R&D)	1-6
A	大野雅之 外 4 名, 分散表現に基づく選択選好モデルの文脈化, 情報処理学会 研究報告 自然言語処理 (NL), 2016-NL-225 [online], 15 January 2016, pp. 1-9 (Research Report of the Information Processing Society of Japan of Natural Language Processing), non-official translation (OONO, Masayuki et al., Contextualization of selective preference model based on distributed representation)	1-6
A	US 2009/0326919 A1 (BEAN, D. L.) 31 December 2009, paragraphs [0278]-[0295] (Family: none)	1-6

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

20 November 2017

Date of mailing of the international search report

05 December 2017

Name and mailing address of the ISA/

Japan Patent Office

3-4-3, Kasumigaseki, Chiyoda-ku,

Tokyo 100-8915, Japan

Authorized officer

Telephone No.

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. G06F17/27(2006.01)i, G06F17/30(2006.01)i

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int.Cl. G06F17/27, G06F17/30

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2017年
日本国実用新案登録公報	1996-2017年
日本国登録実用新案公報	1994-2017年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	江原 暉将 外1名, 確率モデルによる省略主語の補完, NHK技 研R&D, 1997.02.15, 第44号, p.12-28	1-6
A	大野 雅之 外4名, 分散表現に基づく選択選好モデルの文脈化, 情報処理学会 研究報告 自然言語処理 (NL) 2016-NL -225 [online], 2016.01.15, p. 1-9	1-6
A	US 2009/0326919 A1 (BEAN David L.) 2009.12.31, 段落 0278-0295 (ファミリーなし)	1-6

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの
「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
「O」 口頭による開示、使用、展示等に言及する文献
「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」 同一パテントファミリー文献

国際調査を完了した日

20.11.2017

国際調査報告の発送日

05.12.2017

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)

郵便番号 100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

長 由紀子

電話番号 03-3581-1101 内線 3599

5M

4233