

(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
2020年8月13日(13.08.2020)



(10) 国際公開番号

WO 2020/162240 A1

- (51) 国際特許分類:
G06F 40/44 (2020.01) *G10L 15/06* (2013.01)
G10L 15/01 (2013.01) *G10L 15/18* (2013.01)
- (21) 国際出願番号: PCT/JP2020/002650
- (22) 国際出願日: 2020年1月27日(27.01.2020)
- (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 2019-021546 2019年2月8日(08.02.2019) JP
- (71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).
- (72) 発明者: 増村 亮(MASUMURA, Ryo); 〒1808585 東京都武蔵野市緑町三丁目9番11号 N T T 知的財産センタ内 Tokyo (JP). 田中 智大(TANAKA, Tomohiro); 〒1808585 東京都武蔵野市緑町三丁目9番11号 N T T 知的財産センタ内 Tokyo (JP). 大庭 隆伸(OBA, Takanobu); 〒1808585 東京都武蔵野市緑町三丁目9番11号 N T T 知的財産センタ内 Tokyo (JP).
- (74) 代理人: 中尾 直樹, 外 (NAKAO, Naoki et al.); 〒1600022 東京都新宿区新宿三丁目1番22号 新宿NSOビル6階 Tokyo (JP).

(54) Title: LANGUAGE MODEL SCORE CALCULATION DEVICE, LANGUAGE MODEL CREATION DEVICE, METHODS THEREFOR, PROGRAM, AND RECORDING MEDIUM

(54) 発明の名称: 言語モデルスコア計算装置、言語モデル作成装置、それらの方法、プログラム、および記録媒体

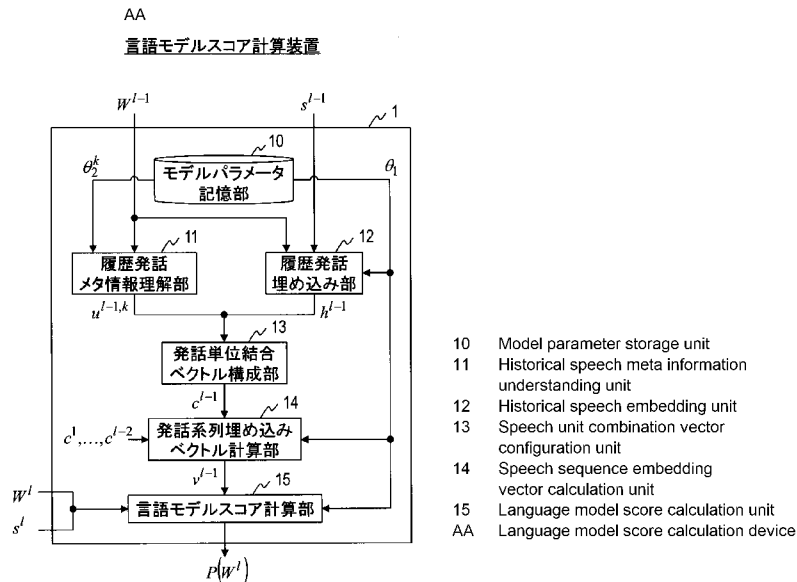


図1

(57) Abstract: The present invention improves the accuracy of language prediction. A historical speech meta information understanding unit 11 derives, from an immediately preceding spoken word string, a historical speech meta information vector, by using a meta information understanding device. A historical speech embedding unit 12 converts the immediately

(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, 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, 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))

preceding spoken word string and a speaker label into a historical speech embedding vector. A speech unit combination vector configuration unit 13 derives a speech unit combination vector by combining the historical speech meta information vector and the historical speech embedding vector. A speech sequence embedding vector calculation unit 14 converts a plurality of speech unit combination vectors derived for past speech sequences into a speech sequence embedding vector. A language model score calculation unit 15 calculates a language model score of current speech from a word string of the current speech, the speaker label, and the speech sequence embedding vector.

(57) 要約: 言語予測の精度を向上する。履歴発話メタ情報理解部 11 は、直前の発話の単語列からメタ情報理解器を用いて履歴発話メタ情報ベクトルを求める。履歴発話埋め込み部 12 は、直前の発話の単語列と話者ラベルとを履歴発話埋め込みベクトルに変換する。発話単位結合ベクトル構成部 13 は、履歴発話メタ情報ベクトルと履歴発話埋め込みベクトルとを結合して発話単位結合ベクトルを求める。発話系列埋め込みベクトル計算部 14 は、過去の発話系列について求めた複数の発話単位結合ベクトルを発話系列埋め込みベクトルに変換する。言語モデルスコア計算部 15 は、現在の発話の単語列と話者ラベルと発話系列埋め込みベクトルとから現在の発話の言語モデルスコアを計算する。

明 細 書

発明の名称：

言語モデルスコア計算装置、言語モデル作成装置、それらの方法、プログラム、および記録媒体

技術分野

[0001] 本発明は、音声認識や機械翻訳等で用いる言語モデルを学習する技術に関する。

背景技術

[0002] 音声認識や機械翻訳では、言語的な予測のために言語モデルが必要である。言語モデルは、言語的な尤もらしさを表す言語モデルスコア（確率）を計測可能なものであり、その性能が音声認識や機械翻訳の性能を左右するものである。これまで、様々な種類の言語モデルが提案されてきているが、近年リカレントニューラルネットワークに基づく言語モデルが注目されている。リカレントニューラルネットワークに基づく言語モデルの詳細は例えば非特許文献1等を参照されたい。

[0003] リカレントニューラルネットワークに基づく言語モデルは、非常に高い言語予測能力を持ち、音声認識や機械翻訳で積極的に利用されている。リカレントニューラルネットワークに基づく言語モデルのポイントは、リカレントニューラルネットワークを用いることにより、当該発話の範囲内の長距離の文脈を反映させた言語予測を行うことができることである。具体的には、ある発話の始端から10番目までの単語の言語的な尤もらしさを計測する際に、始端から直前の9番目までの単語の情報を文脈情報として考慮することができる。

[0004] 一般的なりカレントニューラルネットワークに基づく言語モデルでは、当該発話の範囲内の長距離の文脈情報を考慮して言語モデルスコアを計測するが、会話などでは当該発話の範囲内の文脈情報だけではなく、当該発話よりも過去の発話系列も文脈情報として考慮すべきである。そこで、複数人によ

る会話において当該発話よりも過去の発話系列における話者間のインタラクションまでを文脈情報として考慮して、ある発話の言語モデルスコアを計測する技術が近年検討されている。ここでは、その技術を「複数人会話文脈考慮型言語モデル」と呼ぶこととする。

- [0005] 複数人会話文脈考慮型言語モデルでは、複数人会話において、これまで誰が何を話してきたのかといった情報を陽に活用して、当該発話の単語の予測確率を算出する機能を持ち、一般的なりカレントニューラルネットワークと比較して、予測性能が高い言語モデルスコアを算出することが可能となる。これにより、複数人会話文脈考慮型言語モデルを用いれば、音声認識性能が高い音声認識システムを構築することができる。複数人会話文脈考慮型言語モデルの詳細は非特許文献2を参照されたい。

先行技術文献

非特許文献

- [0006] 非特許文献1: Mikolov Tomas, Karafiat Martin, Burget Lukas, Cernocky Jan, Khudanpur Sanjeev, "Recurrent neural network based language model", INTERSPEECH 2010, pp. 1045-1048, 2010.

非特許文献2: Ryo Masumura, Tomohiro Tanaka, Atsushi Ando, Hirokazu Matsaki, Yushi Aono, "Role Play Dialogue Aware Language Models based on Conditional Hierarchical Recurrent Encoder-Decoder", In Proc. Annual Conference of the International Speech Communication Association (INTERSPEECH), pp. 1259-1263, 2018.

発明の概要

発明が解決しようとする課題

- [0007] 複数人会話文脈考慮型言語モデルでは、当該発話よりも過去の発話系列を文脈情報として利用する際に、各発話の単語列と話者情報のみを利用していた。しかしながら、当該発話にとって文脈情報となり得る、各発話に含まれるメタ情報の理解を明示的に行えていないことによって、不十分な文脈情報

となってしまう。ここで、メタ情報とは、様々なものが考えられるが、例えば「同意」「疑問」「感謝」等を表す発話意図や、「スポーツ」「経済」「芸能」等を表す話題、等が挙げられる。

[0008] 具体的には、例えば当該発話の直前の発話が「そうですか」という場合、この発話が「疑問」を表しているのか「同意」を表しているのかを判断できないと、当該発話でどのような言語が出現しやすいかという言語予測が非常に難しくなる。また、例えば当該発話の直前の発話が「昨日のナイターの試合の結果を教えて」という場合、この発話の話題が「スポーツ」であると具体的に予測できる場合とそうでない場合とでは、次発話の言語予測の精度が大きく異なると考えられる。

[0009] したがって、本発明が解決しようとする課題は、従来の複数人会話文脈考慮型言語モデルが、当該発話の文脈情報となり得る過去の各発話に含まれるメタ情報の理解を明示的に行っていないことにより、言語予測が難しくなっている点である。すなわち、本発明の目的は、過去の発話系列を文脈情報として利用する言語モデルにおいて、言語予測の精度を向上することである。

課題を解決するための手段

[0010] 本発明の第一の態様の言語モデルスコア計算装置は、直前の発話の単語列から少なくとも1個のメタ情報に関するメタ情報理解器を用いて直前の発話のメタ情報を表す履歴発話メタ情報ベクトルを求める履歴発話メタ情報理解部と、直前の発話の単語列と直前の発話の話者を表す話者ラベルとを言語モデルのモデルパラメータを用いて履歴発話埋め込みベクトルに変換する履歴発話埋め込み部と、履歴発話メタ情報ベクトルと履歴発話埋め込みベクトルとを結合して発話単位結合ベクトルを求める発話単位結合ベクトル構成部と、過去の発話系列について求めた複数の発話単位結合ベクトルを言語モデルのモデルパラメータを用いて発話系列埋め込みベクトルに変換する発話系列埋め込みベクトル計算部と、現在の発話の単語列と現在の発話の話者を表す話者ラベルと発話系列埋め込みベクトルとから言語モデルのモデルパラメー

タを用いて現在の発話の言語モデルスコアを計算する言語モデルスコア計算部と、を含む。

[0011] 本発明の第二の態様の言語モデル作成装置は、少なくとも1個のメタ情報に関する発話と各発話のメタ情報との組からなる学習データからメタ情報理解器のモデルパラメータを学習するメタ情報モデルパラメータ学習部と、複数人による複数の発話と各発話の話者を表す話者ラベルとの組からなる会話データからメタ情報理解器のモデルパラメータを用いて言語モデルのモデルパラメータを学習する言語モデルパラメータ学習部と、を含む。

発明の効果

[0012] 本発明によれば、過去の発話系列を文脈情報として利用する言語モデルにおいて、言語予測の精度が向上する。この言語モデルを音声認識や機械翻訳等に用いることによって、音声認識性能や翻訳性能を高めることができる。

図面の簡単な説明

[0013] [図1]図1は、言語モデルスコア計算装置の機能構成を例示する図である。

[図2]図2は、言語モデルスコア計算方法の処理手順を例示する図である。

[図3]図3は、言語モデル作成装置の機能構成を例示する図である。

[図4]図4は、言語モデル作成方法の処理手順を例示する図である。

発明を実施するための形態

[0014] 文中で使用する記号「 $\wedge$ 」は、本来直後の文字の真上に記載されるべきものであるが、テキスト記法の制限により、当該文字の直前に記載する。数式中においてはこれらの記号は本来の位置、すなわち文字の真上に記述している。例えば、「 $\wedge \theta$ 」は数式中では次式で表される。

[0015] [数1]

$$\hat{\theta}$$

[0016] 本発明では、上述の課題を解決するために、複数人会話文脈考慮型言語モデルにおいて、当該発話の文脈情報となる過去の各発話に含まれる1種類以上のメタ情報を推定し、推定した各発話のメタ情報も単語列と話者情報に付

随して文脈情報として用いることによって、言語予測のための予測確率を算出し、それに基づき言語モデルスコアを計算する方法を導入する。そのために、本発明の複数人会話文脈考慮型言語モデルは、推定したいメタ情報の種類数分のメタ情報理解器を内包する。

[0017] また、上述の機能を持つ複数人会話文脈考慮型言語モデル、および内包された推定したいメタ情報の種類数分のメタ情報理解器を、メタ情報の種類ごとのメタ情報予測のための学習データ（発話とメタ情報のペアの複数データ）と、複数人会話の学習データ（話者ラベルが付与された単語列の系列）とを用いて最適化する方法を導入する。

[0018] 以下、本発明の実施の形態について詳細に説明する。なお、図面中において同じ機能を有する構成部には同じ番号を付し、重複説明を省略する。

[0019] [第一実施形態：言語モデルスコア計算装置]

第一実施形態の言語モデルスコア計算装置 1 は、図 1 に例示するように、モデルパラメータ記憶部 10、履歴発話メタ情報理解部 11、履歴発話埋め込み部 12、発話単位結合ベクトル構成部 13、発話系列埋め込みベクトル計算部 14、および言語モデルスコア計算部 15 を備える。この言語モデルスコア計算装置 1 が、図 2 に例示する各ステップの処理を行うことにより第一実施形態の言語モデルスコア計算方法が実現される。

[0020] 言語モデルスコア計算装置 1 は、例えば、中央演算処理装置（CPU: Central Processing Unit）、主記憶装置（RAM: Random Access Memory）などを有する公知又は専用のコンピュータに特別なプログラムが読み込まれて構成された特別な装置である。言語モデルスコア計算装置 1 は、例えば、中央演算処理装置の制御のもとで各処理を実行する。言語モデルスコア計算装置 1 に入力されたデータや各処理で得られたデータは、例えば、主記憶装置に格納され、主記憶装置に格納されたデータは必要に応じて中央演算処理装置へ読み出されて他の処理に利用される。言語モデルスコア計算装置 1 の各処理部は、少なくとも一部が集積回路等のハードウェアによって構成されていてもよい。言語モデルスコア計算装置 1 が備える各記憶部は、例えば、RAM (Rand

om Access Memory) などの主記憶装置、ハードディスクや光ディスクもしくはフラッシュメモリ (Flash Memory) のような半導体メモリ素子により構成される補助記憶装置、またはリレーショナルデータベースやキーバリューストアなどのミドルウェアにより構成することができる。

[0021] 第一実施形態の言語モデルスコア計算装置 1 は、L個の連続した発話ごとの単語列の系列 $W^1, \dots, W^L$ と、L個の連続した発話ごとの話者ラベルの系列 $s^1, \dots, s^L$ とを入力とし、言語モデルのモデルパラメータ θ_1 と、K個のメタ情報理解器のモデルパラメータ $\theta_2^1, \dots, \theta_2^K$ とに従った確率計算により、L個の連続した単語列の系列の言語モデルスコア $P(W^1), \dots, P(W^L)$ を得て出力する。なお、l番目の発話の言語モデルスコア $P(W^l)$ は、厳密には以下を表す。

[0022] [数2]

$$P(W^l) \approx P(W^l | W^1, \dots, W^{l-1}, s^1, \dots, s^l, \theta_1, \theta_2^1, \dots, \theta_2^K)$$

[0023] 言語モデルスコア計算装置 1 は、発話単位の処理を繰り返すことで上記の処理を実現する。以下では、l (l=1, ..., L) 番目の発話単位の処理について詳細を説明する。

[0024] モデルパラメータ記憶部 10 には、言語モデルのモデルパラメータ θ_1 と、K個のメタ情報理解器のモデルパラメータ $\theta_2^1, \dots, \theta_2^K$ とが予め記憶されている。これらのモデルパラメータ $\theta_1, \theta_2^1, \dots, \theta_2^K$ は、後述する言語モデル作成装置 2 により予め学習しておく。

[0025] ステップ S 11 において、履歴発話メタ情報理解部 11 は、l-1番目の発話の単語列 W^{l-1} を入力とし、1以上K以下の各整数kについて、k番目のメタ情報理解器のモデルパラメータ θ_2^k に基づく変換関数により、l-1番目の発話のk番目のメタ情報ベクトル $u^{l-1,k}$ を求める。このとき、l-1番目の発話の単語列 W^{l-1} は1つ以上の単語を含む。履歴発話メタ情報理解部 11 は、得られたK種類のメタ情報ベクトル $u^{l-1,1}, \dots, u^{l-1,K}$ を発話単位結合ベクトル構成部 13 へ出力する。

[0026] メタ情報ベクトルは、単語列から予測されるメタ情報が埋め込まれたベクトルである。メタ情報としては様々なものが考えられるが、「同意」「疑問

」「感謝」などを表す発話意図や、「スポーツ」「経済」「芸能」などを表す話題、等である。メタ情報ベクトルの各次元は、メタ情報に関する各要素に対応させることができる。例えば、1次元目は「発話意図が同意」、2次元目は「発話意図が疑問」などと割り当てておき、例えば、2次元目のみ1、他の次元は0としたベクトルを構成することができる。このとき変換関数はモデルパラメータ θ_2 に基づくメタ情報理解器から抽出可能な情報であれば任意のものをを用いることができる。具体的には、可変長数の記号列を単一ベクトルに変換する関数であれば任意のものを利用できる。

[0027] メタ情報理解器は、単語列からメタ情報予測確率分布を推定するものであれば任意のものをを用いることができ、例えばリカレントニューラルネットワークや双方向リカレントニューラルネットワークとソフトマックス関数を用いて構築することができる。変換関数は、例えばメタ情報理解器が予測したラベルに対応する次元を1として、残りの次元を0としてメタ情報ベクトルを構成する変換関数とすることができる。また、メタ情報理解器が予測したメタ情報についての事後確率分布を直接メタ情報ベクトルとすることができる。さらに、メタ情報理解器を用いて推定する際の間表現を用いてもよく、例えば前述のリカレントニューラルネットワークや双方向リカレントニューラルネットワークとソフトマックス関数を用いたメタ情報理解器の場合であれば、ソフトマックス関数を通す前に得られるベクトル表現を直接メタ情報ベクトルとしてもよい。

[0028] ステップS 12において、履歴発話埋め込み部12は、 $l-1$ 番目の発話の単語列 W^{l-1} と $l-1$ 番目の発話の単語ラベル s^{l-1} とを入力とし、言語モデルのモデルパラメータ θ_1 に基づく変換関数により、 $l-1$ 番目の発話の履歴発話埋め込みベクトル h^{l-1} を求める。履歴発話埋め込みベクトルは、単語列と話者ラベルの情報が埋め込まれたベクトルである。このとき変換関数には、可変長数の記号列を単一ベクトルに変換する関数であれば任意のものを利用できるが、例えばリカレントニューラルネットワークや双方向リカレントニューラルネットワークを用いることができる。履歴発話埋め込み部12は、得られた履歴発

話埋め込みベクトル h^{l-1} を発話単位結合ベクトル構成部13へ出力する。

[0029] ステップS13において、発話単位結合ベクトル構成部13は、履歴発話メタ情報理解部11が出力する $l-1$ 番目の発話の K 種類のメタ情報ベクトル $u^{l-1,1}, \dots, u^{l-1,K}$ と、履歴発話埋め込み部12が出力する $l-1$ 番目の発話の履歴発話埋め込みベクトル h^{l-1} とを入力とし、 $l-1$ 番目の発話の発話単位結合ベクトル c^{l-1} を求める。発話単位結合ベクトル構成部13は、得られた発話単位結合ベクトル c^{l-1} を発話系列埋め込みベクトル計算部14へ出力する。

[0030] 発話単位結合ベクトル c^{l-1} は、次のように構成される。

[0031] [数3]

$$c^{l-1} = [h^{l-1T}, u^{l-1,1T}, \dots, u^{l-1,KT}]^T$$

[0032] ここで、 $\cdot^T$ （上付き添え字の T ）はベクトルの転置を表す。

[0033] ステップS14において、発話系列埋め込みベクトル計算部14は、過去の発話系列について求めた複数の発話単位結合ベクトル $c^1, \dots, c^{l-1}$ を入力とし、言語モデルのモデルパラメータ θ_1 に基づく変換関数により、 $l-1$ 番目の発話系列埋め込みベクトル v^{l-1} を求める。このとき変換関数には、可変長数のベクトル列を単一ベクトルに変換する関数であれば任意のものを利用できるが、例えばリカレントニューラルネットワークを用いることができる。発話系列埋め込みベクトル計算部14は、得られた発話系列埋め込みベクトル v^{l-1} を言語モデルスコア計算部15へ出力する。

[0034] ステップS15において、言語モデルスコア計算部15は、 l 番目の発話の単語列 W^l と l 番目の発話の話者ラベル s^l と $l-1$ 番目の発話系列埋め込みベクトル v^{l-1} とを入力とし、言語モデルのモデルパラメータ θ_1 に基づく変換関数に基づき、 l 番目の発話の言語モデルスコア $P(W^l)$ を求める。このとき変換関数には、自己回帰を行うリカレントニューラルネットワークや双方向リカレントニューラルネットワークとソフトマックス関数を用いて表すことができる。例えば、 l 番目の発話の m 番目の単語と、 l 番目の発話の話者ラベル s^l と、 $l-1$ 番目の発話系列埋め込みベクトル v^{l-1} とから、 $m+1$ 番目の単語を予測する変換関数とすることで、 l 番目の発話 W^l の各単語の確率を求めることができ、それに基づき

、言語モデルスコア $P(W^l)$ を算出することが可能である。

[0035] 言語モデルスコア計算装置 1 は、上述のステップ S 1 1 から S 1 5 までの処理を、入力された L 個の発話 $W^1, \dots, W^L$ それぞれに対して行い、得られた L 個の発話の言語モデルスコア $P(W^1), \dots, P(W^L)$ を出力する。

[0036] [第二実施形態：言語モデル作成装置]

第二実施形態の言語モデル作成装置 2 は、図 3 に例示するように、メタ情報モデルパラメータ学習部 2 1、言語モデルパラメータ学習部 2 2、およびモデルパラメータ記憶部 1 0 を備える。この言語モデル作成装置 2 が、図 4 に例示する各ステップの処理を行うことにより第二実施形態の言語モデル作成方法が実現される。

[0037] 言語モデル作成装置 2 は、例えば、中央演算処理装置 (CPU: Central Processing Unit)、主記憶装置 (RAM: Random Access Memory) などを有する公知又は専用のコンピュータに特別なプログラムが読み込まれて構成された特別な装置である。言語モデル作成装置 2 は、例えば、中央演算処理装置の制御のもとで各処理を実行する。言語モデル作成装置 2 に入力されたデータや各処理で得られたデータは、例えば、主記憶装置に格納され、主記憶装置に格納されたデータは必要に応じて中央演算処理装置へ読み出されて他の処理に利用される。言語モデル作成装置 2 の各処理部は、少なくとも一部が集積回路等のハードウェアによって構成されていてもよい。言語モデル作成装置 2 が備える各記憶部は、例えば、RAM (Random Access Memory) などの主記憶装置、ハードディスクや光ディスクもしくはフラッシュメモリ (Flash Memory) のような半導体メモリ素子により構成される補助記憶装置、またはリレーショナルデータベースやキーバリューストアなどのミドルウェアにより構成することができる。

[0038] 第二実施形態の言語モデル作成装置 2 は、K 個のメタ情報それぞれについての学習データ $D^1, \dots, D^K$ と複数人会話データ D とを入力とし、言語モデルのモデルパラメータ θ_1 と K 個のメタ情報理解器のモデルパラメータ $\theta_2^1, \dots, \theta_2^K$ とを学習する。

[0039] k番目のメタ情報に関する学習データ D^k は、発話 W_n^k とメタ情報 L_n^k との組からなる複数のデータであり、次式で表すことができる。ただし、nは1以上 N^k 以下の各整数であり、 N^k はk番目のメタ情報に関する学習データの数を表す。

[0040] [数4]

$$D^k = \{(W_1^k, L_1^k), \dots, (W_{N^k}^k, L_{N^k}^k)\}$$

[0041] 複数人会話データDは、会話に含まれる発話ごとの単語列 W^l と話者ラベル s^l との組からなる複数のデータであり、次式で表すことができる。ただし、lは1以上L以下の各整数であり、Lは会話に含まれる発話の数を表す。

[0042] [数5]

$$D = \{(W^1, s^1), \dots, (W^L, s^L)\}$$

[0043] ステップS 2 1において、メタ情報モデルパラメータ学習部2 1は、1以上K以下の各整数kについて、k番目のメタ情報に関する学習データ D^k を入力とし、k番目のメタ情報理解器のモデルパラメータ $\hat{\theta}_2^k$ を学習する。学習データ D^k に含まれる発話 W_n^k とメタ情報 L_n^k との組は、例えば1万組準備する（すなわち、 $N^k=10000$ ）。メタ情報モデルパラメータ学習部2 1は、得られたK種類のメタ情報理解器のモデルパラメータ $\hat{\theta}_2^1, \dots, \hat{\theta}_2^K$ を言語モデルパラメータ学習部2 2へ出力する。

[0044] 学習データ D^k により最適化されたモデルパラメータ $\hat{\theta}_2^k$ は次式に従う。

[0045] [数6]

$$\hat{\theta}_2^k = \arg \max_{\theta} \prod_{n=1}^{N^k} P(L_n^k | W_n^k, \theta_2^k)$$

[0046] ここで、 $P(L_n^k | W_n^k, \theta_2^k)$ は、言語モデルスコア計算装置1の履歴発話メタ情報理解部1 1で定義されたメタ情報理解器と同一のものであり、単語列からメタ情報予測確率分布を推定するものであれば任意のものを用いることができる。例えばリカレントニューラルネットワークや双方向リカレントニューラルネットワークとソフトマックス関数を用いて構築することができる。ここで学習されたモデルパラメータ $\hat{\theta}_2^k$ を言語モデルスコア計算装置1における

メタ情報理解器のモデルパラメータ θ_2^k として用いる。

[0047] ステップS 2 2において、言語モデルパラメータ学習部 2 2は、複数人会話データDと、K種類のメタ情報理解器のモデルパラメータ $\hat{\theta}_2^1, \dots, \hat{\theta}_2^K$ とを入力とし、言語モデルのモデルパラメータ $\hat{\theta}_1$ を学習する。複数人会話データDに含まれる発話 W^l と話者ラベル s^l との組は、例えば 1 0 万組準備する（すなわち、 $L=100000$ ）。

[0048] 複数人会話データDにより最適化されたパラメータ $\hat{\theta}_1$ は次式に従う。

[0049] [数7]

$$\begin{aligned}\hat{\theta}_1 &= \arg \max_{\theta_1} \prod_{l=1}^L P(W^l) \\ &= \arg \max_{\theta_1} \prod_{l=1}^L P(W^l | W^1, \dots, W^{l-1}, s^1, \dots, s^l, \theta_1, \hat{\theta}_2^1, \dots, \hat{\theta}_2^K)\end{aligned}$$

[0050] なお、 $P(W^l | W^1, \dots, W^{l-1}, s^1, \dots, s^l, \theta_1, \hat{\theta}_2^1, \dots, \hat{\theta}_2^K)$ は、言語モデルスコア計算装置 1 において定義された計算の流れに従い計算できる。ここで、 θ_1 がここで学習されるモデルパラメータであり、 $\hat{\theta}_2^1, \dots, \hat{\theta}_2^K$ はメタ情報モデルパラメータ学習部 2 1 において最適化されたモデルパラメータである。ここで最適化されたモデルパラメータ $\hat{\theta}_1$ を言語モデルスコア計算装置 1 における言語モデルのモデルパラメータ θ_1 として用いることで、学習した情報を反映した言語モデルスコア計算装置 1 を実現できる。

[0051] 言語モデル作成装置 2 は、上述のステップS 2 1 からS 2 2 までの処理で得られた言語モデルのモデルパラメータ θ_1 とK種類のメタ情報理解器のモデルパラメータ $\hat{\theta}_2^1, \dots, \hat{\theta}_2^K$ とをモデルパラメータ記憶部 1 0 へ記憶する。もしくは、言語モデルおよびメタ情報理解器を用いる他の装置に記憶させるために、得られた言語モデルのモデルパラメータ θ_1 とK種類のメタ情報理解器のモデルパラメータ $\hat{\theta}_2^1, \dots, \hat{\theta}_2^K$ とをそのまま出力する。この場合、言語モデル作成装置 2 はモデルパラメータ記憶部 1 0 を備えなくてもよい。

[0052] [変形例]

上述の実施形態では、言語モデルスコア計算装置 1 と言語モデル作成装置

2とを別々の装置として構成する例を説明したが、言語モデルおよびメタ情報理解器のモデルパラメータを学習する機能と学習済みのモデルパラメータを用いて言語モデルスコアを計算する機能とを兼ね備えた1台の言語モデルスコア計算装置として構成することも可能である。すなわち、変形例の言語モデルスコア計算装置は、メタ情報モデルパラメータ学習部21、言語モデルパラメータ学習部22、モデルパラメータ記憶部10、履歴発話メタ情報理解部11、履歴発話埋め込み部12、発話単位結合ベクトル構成部13、発話系列埋め込みベクトル計算部14、および言語モデルスコア計算部15を備える。

[0053] 以上、本発明の実施の形態について説明したが、具体的な構成は、これらの実施の形態に限られるものではなく、本発明の趣旨を逸脱しない範囲で適宜設計の変更等があっても、本発明に含まれることはいうまでもない。実施の形態において説明した各種の処理は、記載の順に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的にあるいは個別に実行されてもよい。

[0054] [プログラム、記録媒体]

上記実施形態で説明した各装置における各種の処理機能をコンピュータによって実現する場合、各装置が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、上記各装置における各種の処理機能がコンピュータ上で実現される。

[0055] この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよい。

[0056] また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプロ

グラムを転送することにより、このプログラムを流通させる構成としてもよい。

[0057] このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶装置に格納する。そして、処理の実行時、このコンピュータは、自己の記憶装置に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実行形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよく、さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆるASP (Application Service Provider) 型のサービスによって、上述の処理を実行する構成としてもよい。なお、本形態におけるプログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの（コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等）を含むものとする。

[0058] また、この形態では、コンピュータ上で所定のプログラムを実行させることにより、本装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

符号の説明

- [0059] 1 言語モデルスコア計算装置
- 1 0 モデルパラメータ記憶部
 - 1 1 履歴発話メタ情報理解部
 - 1 2 履歴発話埋め込み部
 - 1 3 発話単位結合ベクトル構成部
 - 1 4 発話系列埋め込みベクトル計算部

- 1 5 言語モデルスコア計算部
- 2 言語モデル作成装置
 - 2 1 メタ情報モデルパラメータ学習部
 - 2 2 言語モデルパラメータ学習部

請求の範囲

- [請求項1] 直前の発話の単語列から少なくとも1個のメタ情報に関するメタ情報理解器を用いて上記直前の発話のメタ情報を表す履歴発話メタ情報ベクトルを求める履歴発話メタ情報理解部と、
- 上記直前の発話の単語列と上記直前の発話の話者を表す話者ラベルとを言語モデルのモデルパラメータを用いて履歴発話埋め込みベクトルに変換する履歴発話埋め込み部と、
- 上記履歴発話メタ情報ベクトルと上記履歴発話埋め込みベクトルとを結合して発話単位結合ベクトルを求める発話単位結合ベクトル構成部と、
- 過去の発話系列について求めた複数の発話単位結合ベクトルを上記言語モデルのモデルパラメータを用いて発話系列埋め込みベクトルに変換する発話系列埋め込みベクトル計算部と、
- 現在の発話の単語列と上記現在の発話の話者を表す話者ラベルと上記発話系列埋め込みベクトルとから上記言語モデルのモデルパラメータを用いて上記現在の発話の言語モデルスコアを計算する言語モデルスコア計算部と、
- を含む言語モデルスコア計算装置。
- [請求項2] 少なくとも1個のメタ情報に関する発話と各発話のメタ情報との組からなる学習データからメタ情報理解器のモデルパラメータを学習するメタ情報モデルパラメータ学習部と、
- 複数人による複数の発話と各発話の話者を表す話者ラベルとの組からなる会話データから上記メタ情報理解器のモデルパラメータを用いて言語モデルのモデルパラメータを学習する言語モデルパラメータ学習部と、
- を含む言語モデル作成装置。
- [請求項3] 履歴発話メタ情報理解部が、直前の発話の単語列から少なくとも1個のメタ情報に関するメタ情報理解器を用いて上記直前の発話のメタ

情報を表す履歴発話メタ情報ベクトルを求め、

履歴発話埋め込み部が、上記直前の発話の単語列と上記直前の発話の話者を表す話者ラベルとを言語モデルのモデルパラメータを用いて履歴発話埋め込みベクトルに変換し、

発話単位結合ベクトル構成部が、上記履歴発話メタ情報ベクトルと上記履歴発話埋め込みベクトルとを結合して発話単位結合ベクトルを求め、

発話系列埋め込みベクトル計算部が、過去の発話系列について求めた複数の発話単位結合ベクトルを上記言語モデルのモデルパラメータを用いて発話系列埋め込みベクトルに変換し、

言語モデルスコア計算部が、現在の発話の単語列と上記現在の発話の話者を表す話者ラベルと上記発話系列埋め込みベクトルとから上記言語モデルのモデルパラメータを用いて上記現在の発話の言語モデルスコアを計算する、

言語モデルスコア計算方法。

[請求項4]

メタ情報モデルパラメータ学習部が、少なくとも1個のメタ情報に関する発話と各発話のメタ情報との組からなる学習データからメタ情報理解器のモデルパラメータを学習し、

言語モデルパラメータ学習部が、複数人による複数の発話と各発話の話者を表す話者ラベルとの組からなる会話データから上記メタ情報理解器のモデルパラメータを用いて言語モデルのモデルパラメータを学習する、

言語モデル作成方法。

[請求項5]

請求項1に記載の言語モデルスコア計算装置もしくは請求項2に記載の言語モデル作成装置としてコンピュータを機能させるためのプログラム。

[請求項6]

請求項1に記載の言語モデルスコア計算装置もしくは請求項2に記載の言語モデル作成装置としてコンピュータを機能させるためのプロ

グラムが記録されたコンピュータ読み取り可能な記録媒体。

[図1]

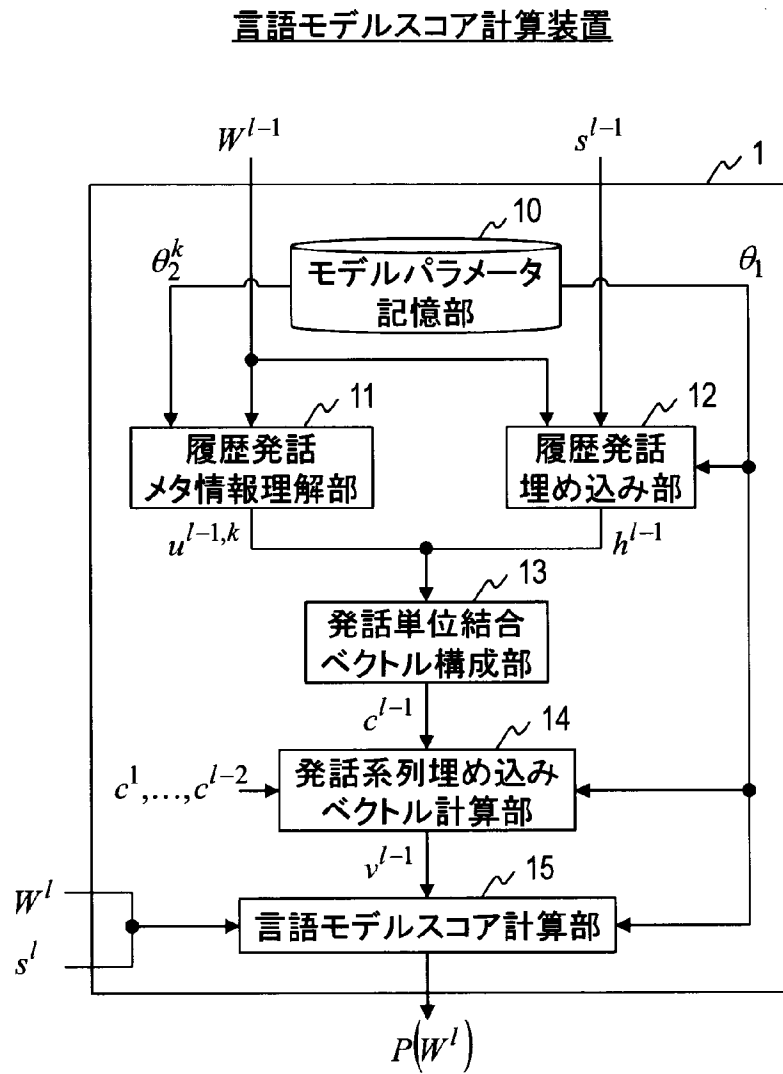


図1

[図2]

言語モデルスコア計算方法

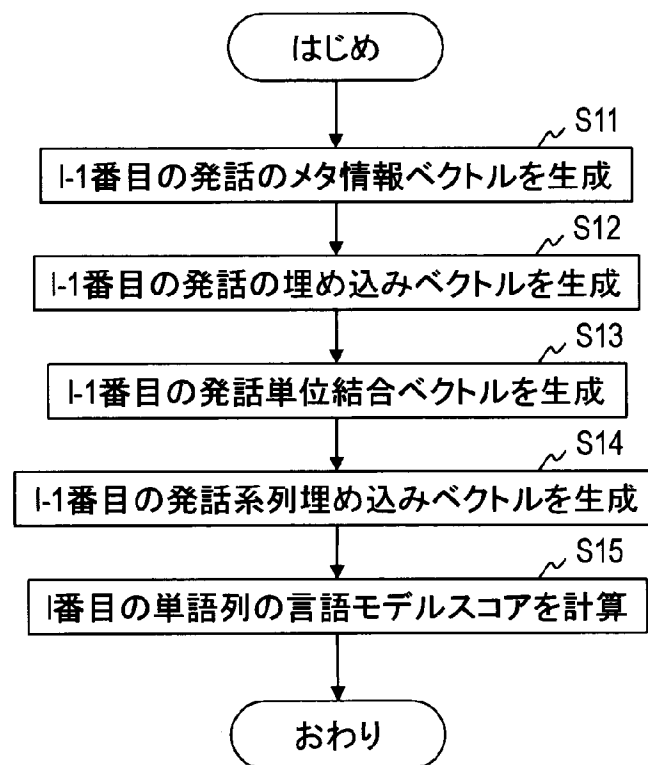


図2

[図3]

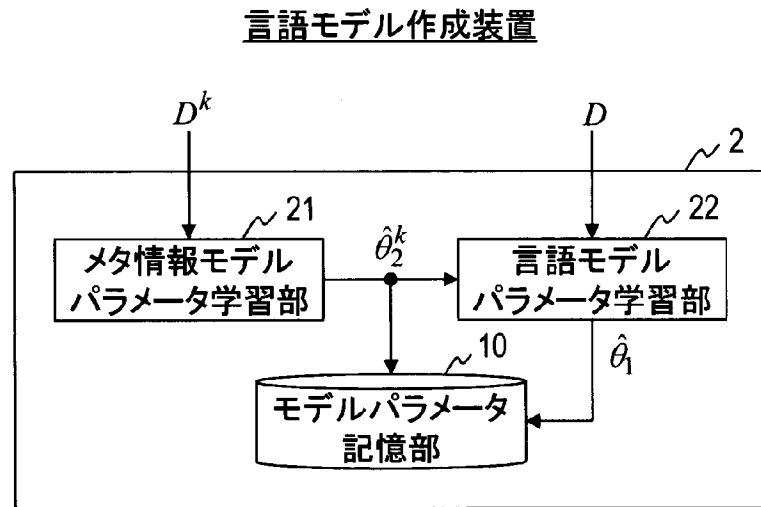


図3

[図4]

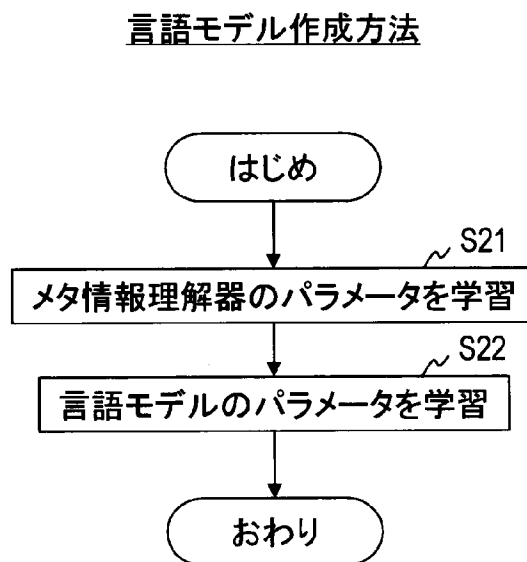


図4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/002650

A. CLASSIFICATION OF SUBJECT MATTER

G06F 40/44 (2020.01)i; G10L 15/01 (2013.01)i; G10L 15/06 (2013.01)i; G10L 15/18 (2013.01)i
FI: G10L15/01 200; G10L15/18 300G; G06F17/28 618; G10L15/06 300C

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)
G10L15/00-15/34; G06F40/44

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

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2020
Registered utility model specifications of Japan	1996-2020
Published registered utility model applications of Japan	1994-2020

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	JP 2018-84627 A (JAPAN BROADCASTING CORPORATION) 31.05.2018 (2018-05-31) paragraphs [0017]-[0023]	1-6
A	WO 2018/051841 A1 (NIPPON TELEGRAPH AND TELEPHONE CORP.) 22.03.2018 (2018-03-22) entire text	1-6
A	US 9984682 B1 (EDUCATIONAL TESTING SERVICE) 29.05.2018 (2018-05-29) entire text	1-6
A	JP 2017-117045 A (NIPPON TELEGRAPH AND TELEPHONE CORP., UNIV. WASEDA) 29.06.2017 (2017-06-29) entire text	1-6
A	JP 2013-117683 A (JAPAN BROADCASTING CORPORATION) 13.06.2013 (2013-06-13) entire text	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
02 April 2020 (02.04.2020)

Date of mailing of the international search report
14 April 2020 (14.04.2020)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/002650

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2014-74732 A (JAPAN BROADCASTING CORPORATION) 24.04.2014 (2014-04-24) entire text	1-6
A	渡部晋治, 堀貴明, 音声言語理解のための音声認識, 電子情報通信学会誌, September 2018, vol. 101, no. 9, ISSN 0913-5693, pp. 885-890, (WATANABE, Shinji, HORI, Takaaki, "Automatic Speech Recognition for Spoken Language Understanding", The journal of the Institute of Electronics, Information and Communication Engineers)	1-6

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/JP2020/002650

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
JP 2018-84627 A	31 May 2018	(Family: none)	
WO 2018/051841 A1	22 Mar. 2018	US 2019/0244604 A1 entire text	
US 9984682 B1	29 May 2018	(Family: none)	
JP 2017-117045 A	29 Jun. 2017	(Family: none)	
JP 2013-117683 A	13 Jun. 2013	(Family: none)	
JP 2014-74732 A	24 Apr. 2014	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06F 40/44(2020.01)i; G10L 15/01(2013.01)i; G10L 15/06(2013.01)i; G10L 15/18(2013.01)i FI: G10L15/01 200; G10L15/18 300G; G06F17/28 618; G10L15/06 300C		
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）） G10L15/00-15/34; G06F40/44 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 1971-2020年 日本国実用新案登録公報 1996-2020年 日本国登録実用新案公報 1994-2020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	JP 2018-84627 A（日本放送協会）31.05.2018（2018-05-31） 段落[0017]-[0023]	1-6
A	WO 2018/051841 A1（日本電信電話株式会社）22.03.2018（2018-03-22） 全文	1-6
A	US 9984682 B1（EDUCATIONAL TESTING SERVICE）29.05.2018（2018-05-29） 全文	1-6
A	JP 2017-117045 A（日本電信電話株式会社，学校法人早稲田大学）29.06.2017 （2017-06-29） 全文	1-6
A	JP 2013-117683 A（日本放送協会）13.06.2013（2013-06-13） 全文	1-6
A	JP 2014-74732 A（日本放送協会）24.04.2014（2014-04-24） 全文	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、一般的な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抵触するものではなく、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 02.04.2020		国際調査報告の発送日 14.04.2020
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号		権限のある職員（特許庁審査官） 山下 剛史 5Z 8946 電話番号 03-3581-1101 内線 3591

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	渡部晋治, 堀貴明, 音声言語理解のための音声認識, 電子情報通信学会誌, 2018.09, 第101巻, 第9号, ISSN 0913-5693 pp. 885-890	1-6

国際調査報告
 パテントファミリーに関する情報

国際出願番号

PCT/JP2020/002650

引用文献			公表日	パテントファミリー文献	公表日
JP	2018-84627	A	31.05.2018	(ファミリーなし)	
WO	2018/051841	A1	22.03.2018	US 2019/0244604 A1	
全文					
US	9984682	B1	29.05.2018	(ファミリーなし)	
JP	2017-117045	A	29.06.2017	(ファミリーなし)	
JP	2013-117683	A	13.06.2013	(ファミリーなし)	
JP	2014-74732	A	24.04.2014	(ファミリーなし)	